

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 11.81 Seconds
(without alignments)
30.703 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASFAIGDTAQPSPV 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCPUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	94.3	185	1 US-08-233-788A-61	Sequence 61, Appl
2	42	48.3	205	1 US-08-277-231A-5	Sequence 5, Appl
3	42	48.3	205	2 US-08-473-750-8	Sequence 8, Appl
4	42	48.3	205	2 US-08-477-326-8	Sequence 8, Appl
5	40	46.0	165	6 5217891-5	Patent No. 5217891
6	40	46.0	1732	2 US-08-570-311-10	Sequence 10, Appl
7	40	46.0	1732	2 US-08-353-485-10	Sequence 10, Appl
8	40	46.0	2628	2 US-08-570-311-14	Sequence 14, Appl
9	39	44.8	387	3 US-08-689-421-23	Sequence 23, Appl
10	39	44.8	387	4 US-09-389-528-23	Sequence 23, Appl
11	39	44.8	387	4 US-09-181-827A-23	Sequence 23, Appl
12	39	44.8	516	4 US-08-689-421-29	Sequence 29, Appl
13	39	44.8	516	4 US-09-389-528-29	Sequence 29, Appl
14	39	44.8	516	4 US-09-181-827A-29	Sequence 29, Appl
15	38	43.7	9	2 US-08-417-174-92	Sequence 92, Appl
16	38	43.7	647	1 US-08-218-943-1	Sequence 1, Appl
17	37	42.5	1876	2 US-08-619-554-2	Sequence 2, Appl
18	37	42.5	1895	2 US-08-619-554-4	Sequence 4, Appl
19	36	41.4	245	2 US-08-272-255-15	Sequence 15, Appl
20	36	41.4	245	5 PCP-US95-08565-15	Sequence 15, Appl
21	36	41.4	574	6 5378464-3	Patent No. 5378464
22	36	41.4	681	2 US-08-272-255-6	Sequence 6, Appl
23	36	41.4	681	2 US-08-272-255-9	Sequence 9, Appl
24	36	41.4	681	4 US-08-964-268-5	Sequence 5, Appl
25	36	41.4	681	4 PCP-US95-08565-6	Sequence 6, Appl
26	36	41.4	681	5 PCP-US95-08565-9	Sequence 9, Appl
27	36	41.4	808	3 US-08-804-439A-14	Sequence 14, Appl

28	36	41.4	808	3 US-08-720-229-14	Sequence 14, Appl
29	36	41.4	830	1 US-08-110-158-4	Sequence 4, Appl
30	36	41.4	830	5 PCT-US91-05059-2	Sequence 2, Appl
31	35	40.2	9	2 US-08-417-174-04	Sequence 94, Appl
32	35	40.2	105	2 US-08-659-251-9	Sequence 9, Appl
33	35	40.2	105	2 US-09-256-490-9	Sequence 9, Appl
34	35	40.2	105	5 PCT-US96-11445-9	Sequence 9, Appl
35	35	40.2	467	2 US-08-727-548-2	Sequence 2, Appl
36	35	40.2	855	2 US-08-619-554-6	Sequence 6, Appl
37	35	40.2	874	3 US-08-804-439A-16	Sequence 16, Appl
38	35	40.2	874	3 US-08-720-229-16	Sequence 2, Appl
39	34	39.1	507	2 US-08-409-122-2	Sequence 2, Appl
40	34	39.1	507	2 US-08-408-669-2	Sequence 2, Appl
41	34	39.1	720	4 US-09-000-016-7	Sequence 7, Appl
42	34	39.1	734	4 US-09-000-016-4	Sequence 2, Appl
43	34	39.1	823	4 US-09-000-016-2	Sequence 2, Appl
44	34	39.1	874	3 US-08-804-439A-15	Sequence 15, Appl
45	34	39.1	874	3 US-08-720-229-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-61
Sequence 61, Application US/0823788A
Patent No. 5635617-2
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESSES:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-6031
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERY
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 61:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-61

Print

Query Match 94.3%; Score 82; DB 1; Length 185;
Best Local Similarity 94.4%; Pred. No. 5.8e-07;
Matches 17; Conservative 1; Mismatches 0; Gaps 0;
QY 1 ASFAIGDTAQPSPV 18

DB 65 ASFTAGTAAVPSIV 82

RESULT 2

US-08-277-231A-5
Sequence 5, Application US/08277231A
Patent No. 5643725

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5643725

TITLE OF INVENTION: Structural Genes and The LKP Pili Operon of No. 5643725typab
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive

CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,231A
FILING DATE: 19-JUL-1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 32,542

REFERENCE/DOCKET NUMBER: ACC94-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-277-231A-5

Query Match 48.3%; Score 42; DB 1; Length 205;

Best Local Similarity 52.9%; Pred. No. 5.3;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGTAAVPSI 17

DB 96 ANLSHAGTAAVPSI 112

RESULT 3

US-08-473-750-8
Sequence 8, Application US/08473750
Patent No. 5834187
Patent No. 5834187 5786143

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5834187

TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5834187 5786143

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,750
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,321

FILING DATE: 19-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.

REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02B
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 205 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-473-750-8

Query Match 48.3%; Score 42; DB 2; Length 205;

Best Local Similarity 52.9%; Pred. No. 5.3;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGTAAVPSI 17

DB 96 ANLSHAGTAAVPSI 112

RESULT 4

US-08-477-326-8
Sequence 8, Application US/08477326
Patent No. 5968769

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5968769

TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5968769typab
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive

CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,326
FILING DATE: 07-JUN-1995

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/277,231

FILING DATE: July 19, 1994

Thu Jun 21 08:38:31 2001

us-09-646-043-1.rai

Page 3

ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-326-8

Query Match 48.3%; Score 42; DB 2; Length 205;
Best Local Similarity 52.9%; Pred. No. 5.3;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 ASPTAGDTTAQYPSI 17
DB 96 ANLSHAGQTAAPVPSI 112

RESULT 5
5217891-5
PATENT NO. 5217891
APPLICANT: BRAKE, ANTHONY J.; VAN DEN BERG, JOHAN A.
TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYEROMYCES
A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
POLYPEPTIDES
NUMBER OF SEQUENCES: 23
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,398
FILING DATE: 09-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 78,551
FILING DATE: 28-JUL-1987
SEQ ID NO: 5
LENGTH: 165
5217891-5

Query Match 46.0%; Score 40; DB 6; Length 165;
Best Local Similarity 53.3%; Pred. No. 9.1;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 TAIGDTTAQYPSI 18
DB 24 TTTGDTTAQIPAEV 38

RESULT 6
US-08-570-311-10
SEQUENCE 10, Application US/08570311
GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepline, Guyalaine
APPLICANT: Han, Naiming
APPLICANT: Lantzi, Joseph
APPLICANT: Patti, Marilyn
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: DP15-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1732 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-10

Query Match 46.0%; Score 40; DB 2; Length 1732;
Best Local Similarity 37.5%; Pred. No. 1.3e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 SPTAGDTTAQYPSI 17
DB 1658 NITSADVTAKKPYTL 1673

RESULT 7
US-08-353-485-10
SEQUENCE 10, Application US/08353485
GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepline, Guyalaine
APPLICANT: Han, Naiming
APPLICANT: Lantzi, Joseph
APPLICANT: Patti, Marilyn
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UP15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1732 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-485-10

Query Match
Best Local Similarity 46.0%; Score 40; DB 2; Length 1732;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFTAGDTAQQVPSI 17
:|::|1111::
Db 1658 NITSLADVTAKRYTL 1673

RESULT 8
US-08-570-311-14
Sequence 14, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UP15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-14

Query Match
Best Local Similarity 46.0%; Score 40; DB 2; Length 2628;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFTAGDTAQQVPSI 17
:|::|1111::
Db 2554 NITSLADVTAKRYTL 2569

RESULT 9
US-08-689-421-23
Sequence 23, Application US/08689421
Patent No. 6008029
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberley M.
APPLICANT: Kauppinen, Sakari
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6008029disk of No. 6008029th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4554.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-689-421-23

Query Match

44.8%; Score 39; DB 3; Length 387;

Thu Jun 21 08:38:31 2001

US-09-646-043-1.rai

Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPS 16
1: 11:1111
Db 222 AVDPSTSQLPFS 233

RESULT 10
US-09-389-528-23
Sequence 23, Application US/09389528
Patent No. 6207430

GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberly M.
APPLICANT: Kauppinen, Sakari P
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESS: No. 62074300, No. 62074300disk of No. 6207430th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/389,528
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-389-528-23

Query Match 44.8%; Score 39; DB 4; Length 387;
Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPS 16
1: 11:1111
Db 222 AVDPSTSQLPFS 233

RESULT 11
US-09-181-827A-23
Sequence 23, Application US/09181827A
Patent No. 6242232

GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberly M.
APPLICANT: Kauppinen, Sakari P
APPLICANT: Halkier, Torben P.

TITLE OF INVENTION: Purified Coprinus Laccases And Nucleic
Acids Encoding Same
FILE REFERENCE: 4554,200-US
CURRENT APPLICATION NUMBER: US/09/181,827A
CURRENT FILING DATE: 1998-10-28
PRIOR FILING DATE: 1995-08-25
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 387
TYPE: PRT
ORGANISM: Coprinus cinereus

US-09-181-827A-23

Query Match 44.8%; Score 39; DB 4; Length 387;
Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPS 16
1: 11:1111
Db 222 AVDPSTSQLPFS 233

RESULT 12
US-08-689-421-29
Sequence 29, Application US/08689421
Patent No. 6008029

GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberly M.
APPLICANT: Kauppinen, Sakari P
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESS: No. 60080290, No. 60080290disk of No. 6008029th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-689-421-29

Query Match 44.8%; Score 39; DB 3; Length 516;
Best Local Similarity 58.3%; Pred. No. 50;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16
1: 11:1111
Db 306 AVDPPTSOLPFS 317

RESULT 13

US-09-389-528-29
Sequence 29, Application US/09389528
Patent No. 6207430
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberley M.
APPLICANT: Kauppinen, Sakari
APPLICANT: Halkier, Torben P.
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 62074300 No. 6207430disk of No. 6207430th America, Inc.
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/389,528
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4554, 204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-389-528-29

Query Match

Best Local Similarity 44.8%; Score 39; DB 4; Length 516;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16
1: 11:1111
Db 306 AVDPPTSOLPFS 317

RESULT 14

US-09-181-827A-29
Sequence 29, Application US/09181827A
Patent No. 6242232
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberley M.
APPLICANT: Kauppinen, Sakari
APPLICANT: Halkier, Torben P.
TITLE OF INVENTION: Purified Coprinus Laccases And Nucleic
FILE REFERENCE: 4554, 200-US

;; CURRENT APPLICATION NUMBER: US/09/181,827A
;; CURRENT FILING DATE: 1998-10-28
;; PRIOR APPLICATION NUMBER: 60/002,800
;; PRIOR FILING DATE: 1995-08-25
;; NUMBER OF SEQ ID NOS: 36
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 29
;; LENGTH: 516
;; TYPE: PRT
;; ORGANISM: Coprinus cinereus
US-09-181-827A-29

Query Match
Best Local Similarity 44.8%; Score 39; DB 4; Length 516;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16
1: 11:1111
Db 306 AVDPPTSOLPFS 317

RESULT 15

US-08-417-174-92
Sequence 92, Application US/08417174
Patent No. 5844075
GENERAL INFORMATION:
APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,
TITLE OF INVENTION: MELANOMA ANTIGENS AND
TITLE OF INVENTION: THEIR USE IN DIAGNOSTIC AND THERAPEUTIC
NUMBER OF SEQUENCES: 126
CORRESPONDENCE ADDRESS:
ADDRESSER: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,174
FILING DATE: 05-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,565
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPPI
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4124US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 9
TYPE: amino acid
STRANDEDNESS: Unknown
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide
US-08-417-174-92

Query Match
Best Local Similarity 43.7%; Score 38; DB 2; Length 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Thu Jun 21 08:38:31 2001

us-09-646-043-1.ral

Page 7

OY 10 TAOVPSI 17
|||||:
Db 2 TAOVPSV 9

Search completed: June 20, 2001, 13:31:40
Job time: 18 sec

Thu Jun 21 08:38:33 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 13.45 Seconds
(without alignments)
101.944 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87
Sequence: 1 ASFTAGDTTAQVPSIV 18

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR.68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	87	100.0	184 2 S20682	type 1 fimbrial pr
2	82	94.3	185 2 B28393	type 1 fimbrial pr
3	49	56.3	191 2 A64785	probable fimbrial-ABC-transporter St
4	49	56.3	191 2 C85553	fimbrial protein h
5	47	54.0	171 2 B64785	probable fimbrial-ABC-transporter St
6	44.5	51.1	591 2 S65588	probable fimbrial-ABC-transporter St
7	44	50.6	171 2 C85553	probable fimbrial-ABC-transporter St
8	44	50.6	931 2 B85052	probable fimbrial-ABC-transporter St
9	43	49.4	216 2 S54430	hypothetical prote
10	43	49.4	659 2 B96827	hypothetical prote
11	43	49.4	1248 2 B96827	hypothetical prote
12	42	48.3	194 2 C85976	probable tyrosine-hypothetical prote
13	42	48.3	445 2 T41416	hypothetical prote
14	42	48.3	1146 2 E70204	long polar fimbria
15	42	47.1	177 2 A56271	conserved hypochet
16	41	47.1	199 2 A81725	fimbrial protein - Esc
17	41	47.1	215 2 S56540	fimbrial protein - Esc
18	41	47.1	215 2 S56540	fimbrial protein - Esc
19	41	47.1	504 2 P1W31	HLI protein human
20	41	47.1	628 2 F84219	DNA-binding protei
21	41	47.1	1298 2 T47523	fimbrial morpholog
22	40	46.0	176 2 S56543	fimbrial morpholog
23	40	46.0	176 2 H86129	hypothetical prote
24	40	46.0	1419 2 T32970	lysine-specific cy
25	40	46.0	1732 2 T30836	hemagglutinin A -
26	40	46.0	2628 2 T28651	hypothetical prote
27	39	44.8	270 2 T43039	hypothetical prote
28	39	44.8	409 2 S77262	hypothetical prote
29	39	44.8	468 2 B70932	probable PPE prote

us-09-646-043-1.rpr

30	39	44.8	488 1 QOBER5	alkaline exonuclease
31	39	44.8	488 2 T44030	alkaline exonuclease
32	39	44.8	488 2 T44215	alkaline exonuclease
33	39	44.8	615 2 T39758	proline-serine ric
34	39	44.8	725 2 E86790	hypothetical prote
35	38	43.7	85 2 T17696	hypothetical prote
36	38	43.7	170 2 T39892	hypothetical prote
37	38	43.7	175 2 S06193	stap protein - Esc
38	38	43.7	175 2 S06193	stap protein - Esc
39	38	43.7	175 2 S06193	stap protein - Esc
40	38	43.7	175 2 S06193	stap protein - Esc
41	38	43.7	202 2 C71306	probable major fimb
42	38	43.7	218 2 F75404	ABC transporter, p
43	38	43.7	259 2 T23782	hypothetical prote
44	38	43.7	347 2 S43771	phosphatidylcholin
45	38	43.7	458 2 T27217	hypothetical prote
	38	43.7	460 2 A84281	hypothetical prote

ALIGNMENTS

RESULT 1
S20682 type 1 fimbrial protein fima - Salmonella typhi
C:Species: Salmonella typhi
C>Date: 20-Feb-1995 #sequence-revision 20-Feb-1995 #text-change 26-Aug-1999
C:Accession: S20682
R:Rossolini, G.M.; Muscas, P.; Chlesurini, A.; Sakta, G.
submitted to the EMBL Data Library, March 1992
A:Reference number: S20682
A:Accession: S20682
A:Molecule type: DNA
A:Residues: 1-184 <ROS>
A:Cross-references: EMBL:X65168, NID:g395382, PIDN:CAA46286.1; PID:g47667
A:Experimental source: strain Scy4
A:Gene: fima
C:Superfamily: type 1 fimbrial protein
C:Keywords: fimbria

Query Match 100.0%; Score 87; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 8.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18
DB 65 ASFTAGDTTAQVPSIV 82

RESULT 2
B28393 type 1 fimbrial protein precursor - Salmonella typhimurium
C:Species: Salmonella typhimurium
C>Date: 28-Aug-1989 #sequence-revision 28-Aug-1989 #text-change 15-Nov-1996
C:Accession: B28393; A05121
R:Purcell, B.R.; Pruckner, J.; Clegg, S.
J. Bacteriol. 169, 3631-3634, 1987
A:Title: Nucleotide sequences of the genes encoding type 1 fimbrial subunits of Klebs
A:Reference number: A91858; M01D:88058806
A:Accession: B28393
A:Molecule type: DNA
A:Residues: 1-185 <PUN>
A:Note: The authors translated the codon GCC for residue 18 as Gly, GTG for residue 3
R:Maalen, K.; Stetten, K.; Froholm, L.O.; Valsanen, V.; Korhonen, T.K.
FEMS Microbiol. Lett. 16, 149-151, 1983
A:Reference number: A05121
A:Accession: A05121
A:Molecule type: protein
A:Residues: 23-30, 'S', 32-45, 'X', 47-50 <MAA>
C:Superfamily: type 1 fimbrial protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-185/Product: type 1 fimbrial protein #status predicted <MAT>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 8.38 Seconds
(without alignments)
73.580 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASFTAIGDTTAQVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	184	1 FMA_SALTY	P37920 salmonella
2	82	94.3	185	1 FMT_SALTY	P55723 salmonella
3	82	94.3	185	1 FMA_SALTY	P37921 salmonella
4	49	56.3	191	1 SFMA_ECOLI	P77660 escherichia
5	47	54.0	171	1 SFMF_ECOLI	P38052 escherichia
6	43	49.4	216	1 HFD1_HAEIN	P45992 haemophilus
7	42	48.3	205	1 HFD2_HAEIN	P45993 haemophilus
8	41	47.1	178	1 LPFA_SALTY	P33660 salmonella
9	41	47.1	211	1 VPV_BPP2	P31340 bacteriophage
10	41	47.1	215	1 FIM1_ECOLI	P39284 escherichia
11	41	47.1	504	1 VLI_HPV31	P17388 human papill
12	40	46.0	176	1 FIMF_ECOLI	P08189 escherichia
13	40	46.0	2628	1 HAGA_PORGI	P51845 porphyromon
14	39	44.8	488	1 EXON_HSV6U	P24447 human herpe
15	39	44.8	488	1 EXON_HSV62	P52448 human herpe
16	38	43.7	170	1 SFAG_ECOLI	P42961 bacillus su
17	38	43.7	175	1 SFAG_ECOLI	P13429 escherichia
18	38	43.7	193	1 ISP2_VITS1	Q9X50 vitreoscilli
19	38	43.7	1135	1 VGLM_HANTV	P08668 hantaan vir
20	38	43.7	1246	1 VP03_HSVSA	Q01000 herpesvirus
21	38	43.7	1501	1 SN02_YEAST	P32568 saccharomyc
22	37	42.5	227	1 PGSA_MYCPN	P75520 mycoplasma
23	37	42.5	285	1 GLO4_YEAST	Q12320 saccharomyc
24	37	42.5	354	1 YAUG_SCHPO	Q10169 schizosacch
25	37	42.5	367	1 FLGI_PSEPU	Q10366 schizosacch
26	37	42.5	851	1 YDBG_SCHPO	Q14053 schizosacch
27	37	42.5	902	1 YC47_SCHPO	P78595 candida alb
28	37	42.5	1499	1 CDR2_CANAL	P38631 saccharomyc
29	37	42.5	1876	1 GLS1_YEAST	P40989 saccharomyc
30	37	42.5	1895	1 GLS2_YEAST	P43215 phleum prat
31	36	41.4	132	1 MPP6_PHLPR	P87072 neurospora
32	36	41.4	174	1 CALB_NEUCR	P42913 escherichia
33	36	41.4	194	1 YRAH_ECOLI	

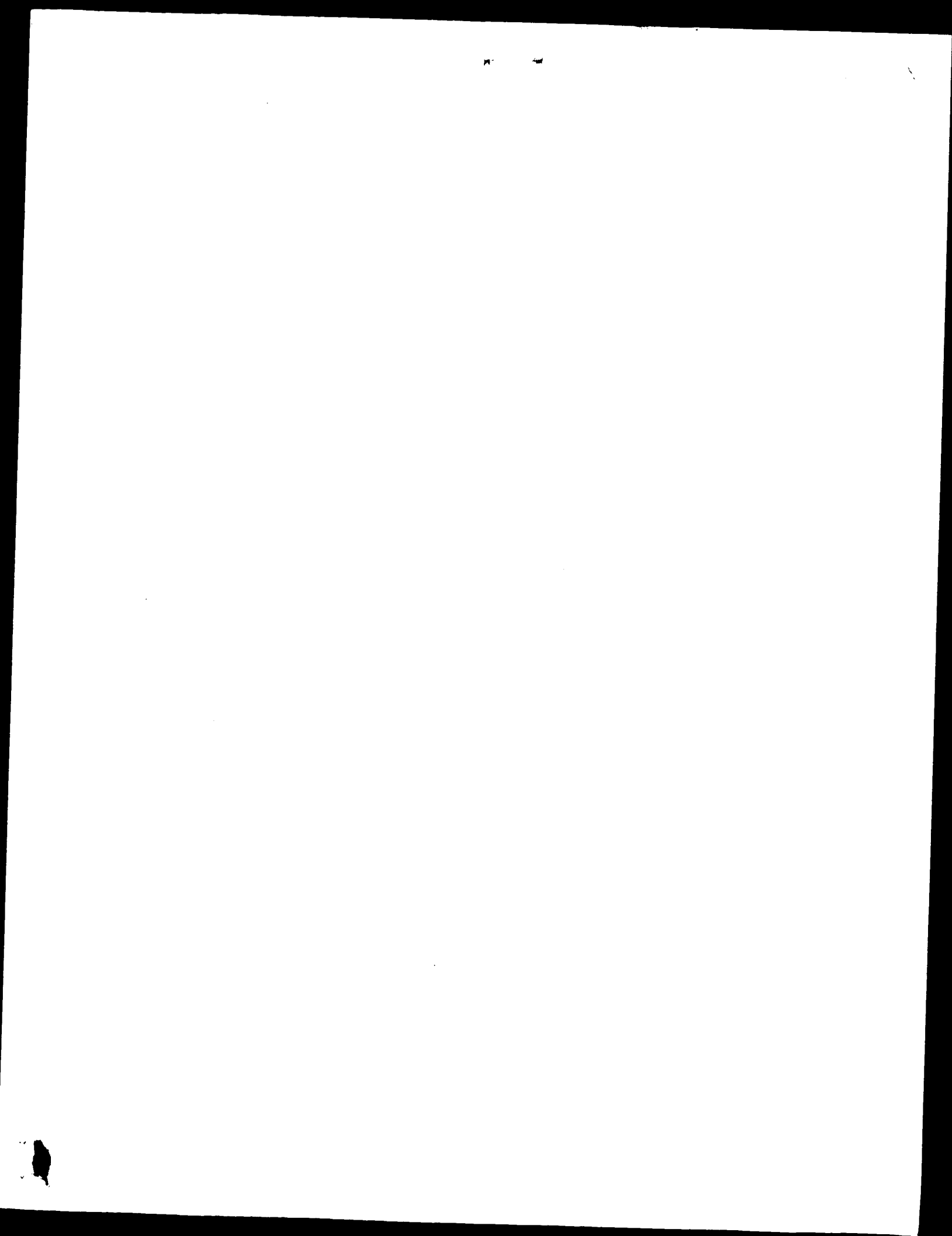
34	36	41.4	198	1 YAEH_HAEIN	P46492 haemophilus
35	36	41.4	269	1 YOEI_ECOLI	Q46942 escherichia
36	36	41.4	278	1 HT22_MYCPN	P46604 arabidopsis
37	36	41.4	312	1 LDH_MYCPN	P78007 mycoplasma
38	36	41.4	393	1 TRUA_MOUSE	Q9WU56 mus musculu
39	36	41.4	444	1 SLAP_LACAC	P35829 lactobacill
40	36	41.4	530	1 TP6B_SULSH	O05207 sulfolobus
41	36	41.4	576	1 Y134_HUMAN	Q14147 homo sapien
42	36	41.4	681	1 CRY1_ARATH	Q43125 arabidopsis
43	36	41.4	808	1 VGLB_HSVSA	P24905 herpesvirus
44	36	41.4	830	1 LEM3_HUMAN	P16109 homo sapien
45	36	41.4	925	1 PIP1_YEAST	P40020 saccharomyc

ALIGNMENTS

RESULT	1	FMIA_SALTY	STANDARD;	PRT;	184 AA.
ID	FMIA_SALTY				
AC	P37920;				
DT	01-OCT-1994 (Rel. 30, Created)				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).				
GN	FIMA.				
OS	Salmonella typhi.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_TaxID=601;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-STVA;				
RA	MEDLINE=94116931; PubMed=8288102;				
RR	Rossolini G.M., Muscas P., Chiesurin A., Satta G.;				
RT	"Analysis of the Salmonella fim gene cluster: identification of a new				
RT	gene (fimi) encoding a fimbria-like protein and located downstream				
RT	from the fima gene".				
RL	FEMS Microbiol. Lett. 114:259-266(1993);				
CC	FUNCTION: FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).				
CC	FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5				
CC	MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO				
CC	COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.				
CC	-1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL; X65168; CAA46286.1; ..				
DR	PIR; S20682; S20682.				
DR	InterPro; IPR000259; ..				
DR	Pfam; PF00419; Fimbrin; 1.				
KW	Fimbria; Signal.				
FT	SIGNAL 1 22				
FT	CHAIN 23 184				
FT	DISULFID 46 86				
FT	SEQUENCE 184 AA; 18793 MW; E46BAD1A22468EB0 CRC64;				
SQ	POTENTIAL.				
	TYPE-1 FIMBRIAL PROTEIN, A CHAIN.				
	PROBABLE.				

Query Match 100.0%; Score 87; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 3.6e-08; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 ASFTAIGDTTAQVPSIV 18
Db 65 ASFTAIGDTTAQVPSIV 82



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 8.38 Seconds

(without alignments)

73.560 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASPTAIGDTTAQVPSIV 18

Scoring table: BLOSUM62

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	184	1 FMA_SALTI	P37920 salmonella
2	82	94.3	185	1 FMA_SALTI	P37923 salmonella
3	82	94.3	185	1 FMA_SALTI	P37921 salmonella
4	49	56.3	191	1 FMA_SALTI	P77660 escherichia
5	47	54.0	171	1 FMA_SALTI	P38052 escherichia
6	43	49.4	216	1 HFD1_HAEIN	P45992 haemophilus
7	42	48.3	205	1 HFD2_HAEIN	P45993 haemophilus
8	41	47.1	178	1 LPPA_SALTY	P43660 salmonella
9	41	47.1	211	1 VPI_BPP2	P31340 bacteriophage
10	41	47.1	215	1 FIM1_ECOLI	P33264 escherichia
11	41	47.1	215	1 FIM1_ECOLI	P17388 human papill
12	40	46.0	176	1 FIME_ECOLI	P08189 escherichia
13	40	46.0	2628	1 HAGA_FORGI	P051845 porphyromon
14	39	44.8	488	1 EXON_HSV62	P24447 human herpes
15	39	44.8	488	1 EXON_HSV62	P52448 human herpes
16	38	43.7	170	1 YCSD_SACSU	P42961 bacillus su
17	38	43.7	175	1 YCSD_SACSU	P13429 escherichia
18	38	43.7	193	1 ISPE_VITSL	O9x550 vitreoscilla
19	38	43.7	135	1 VGLM_HANTV	P08668 hantaan vir
20	38	43.7	1246	1 VP03_HSVSA	001000 herpesvirus
21	38	43.7	1501	1 SMO2_YEAST	P32568 saccharomyc
22	37	42.5	227	1 PGSA_MYCEN	P75520 mycoplasma
23	37	42.5	285	1 GLO4_YEAST	O12320 schistosom
24	37	42.5	354	1 YAOG_SCHPO	O510169 schistosom
25	37	42.5	367	1 YAOG_SCHPO	O510169 schistosom
26	37	42.5	851	1 YDBG_SCHPO	O510169 schistosom
27	37	42.5	902	1 YCAF_SCHPO	O510169 schistosom
28	37	42.5	1499	1 CDR2_CANTL	O14036 schistosom
29	37	42.5	1876	1 GUS2_YEAST	P78593 candida alb
30	37	42.5	1895	1 GUS2_YEAST	P38631 saccharomyc
31	36	41.4	132	1 MPP6_PHLPR	P40988 saccharomyc
32	36	41.4	174	1 CALB_NEUCR	P43215 phleum prat
33	36	41.4	194	1 YRAH_ECOLI	P87072 neuropeptid

34	36	41.4	198	1 YAEI_HAEIN	P46492 haemophilus
35	36	41.4	269	1 YOEI_ECOLI	O46942 escherichia
36	36	41.4	278	1 HT22_ARATH	P46604 arabidopsis
37	36	41.4	312	1 LDH_MYCN	P78007 mycoplasma
38	36	41.4	393	1 TRUA_MOUSE	O9w456 mus muscu
39	36	41.4	444	1 SLAP_LACAC	P35829 lactobacill
40	36	41.4	530	1 TP6B_SULSH	O05207 sulfolobus
41	36	41.4	576	1 Y134_HUMAN	O14147 homo sapien
42	36	41.4	681	1 CRY1_ARATH	O43125 arabidopsis
43	36	41.4	808	1 VGLB_HSVSA	P24905 herpesvirus
44	36	41.4	830	1 LEM3_HUMAN	P16109 homo sapien
45	36	41.4	925	1 PIP1_YEAST	P40020 saccharomyc

ALIGNMENTS

RESULT 1
ID FMA_SALTI STANDARD; PRT; 184 AA.
AC P37920;
DT 01-OCT-1994 (Rel. 30, Created)
DR 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).
CN FIMA.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STY4;
RA MEDLINE-94116831; PubMed-8288102;
RX Rossolini G.M., Muscas P., Chiesurin A., Satta G.;
RT "Analysis of the salmonella fim gene cluster: identification of a new
RT gene (fimb) encoding a fimbriin-like protein and located downstream
RT from the fimb gene."
RL FEBS Microbiol. Lett. 114:259-266(1993)
CC FUNCTION: FIMBRIN (ALSO CALLED PILIN). POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL; X65168; CAA46286.1; -
DR PIR; S20682; S20682.
DR InterPro: IPR000259; -
DR Pfam: PF00419; Fimbrin; 1.
KW Fimbrin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 184
FT DISULFID 46 86
SQ SEQUENCE 184 AA; 18793 MW; E46BAD1A22468EB0 CMC64;
POTENTIAL.
TYPE-1 FIMBRIAL PROTEIN, A CHAIN.
PROBABLY.

Query Match 100.0%; Score 87; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 ASPTAIGDTTAQVPSIV 18
65 ASPTAIGDTTAQVPSIV 82

```

RESULT 2
FM1_SALTY STANDARD; PRT; 185 AA.
AC P55223;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FIMBRIAL SUBUNIT TYPE 1 PRECURSOR.
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8058806; PubMed=2890624;
RA Purcell B.K., Pruckler J., Clegg S.;
RT "Nucleotide sequences of the genes encoding type 1 fimbrial subunits
of Klebsiella pneumoniae and Salmonella typhimurium."
RL J. Bacteriol. 169:5831-5834(1987).
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: WITH E.COLI AND K.PNEUMONIAE FIMBRIAL SUBUNITS TYPE 1
CC PRECURSORS.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: M18283; AAA27063.1; -
DR InterPro: IPR000259; -
DR Pfam: PF00419; Fimbrial; 1.
KW Fimbria; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 185 FIMBRIAL SUBUNIT TYPE 1.
FT DISULFID 46 86 PROBABLE.
SQ SEQUENCE 185 AA; 18897 MW; 52B73721952CB793 CRC64;

Query Match 94.3%; Score 82; DB 1; Length 185;
Best Local Similarity 94.4%; Pred. No. 2.7e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18
Db 65 ASFTAGDTTAQVPSIV 82

RESULT 3
FM1_SALTY STANDARD; PRT; 185 AA.
AC P37921;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).
GN FIMA.
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA Swenson D.L., Clegg S.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
SQ SEQUENCE OF 23-50.
STRAIN=LT2 / SH6749;
Paalen K., Stetten K., Froeholm L.O., Vaeisaenen V., Korhonen T.K.;

```

```

RT "The N-terminal amino acid sequence of type 1 fimbria (pil1) of
RT Salmonella typhimurium LT2."
RL FEBS Microbiol. Lett. 16:149-151(1983).
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: L19338; AAA75416.1; -
DR StyGene: SG10275; fima.
DR InterPro: IPR000259; -
DR Pfam: PF00419; Fimbrial; 1.
KW Fimbria; Signal.
FT SIGNAL 1 22 TYPE-1 FIMBRIAL PROTEIN, A CHAIN.
FT CHAIN 23 185 PROBABLE.
FT DISULFID 46 86
SQ SEQUENCE 185 AA; 18897 MW; 7C9C6E926AD11EB2 CRC64;

Query Match 94.3%; Score 82; DB 1; Length 185;
Best Local Similarity 94.4%; Pred. No. 2.7e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18
Db 65 ASFTAGDTTAQVPSIV 82

RESULT 4
SFMA_ECOLI STANDARD; PRT; 191 AA.
ID SFMA_ECOLI
AC P77660;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SFM FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).
GN SFMA.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MC1655;
RX MEDLINE=9742617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kallman S., Komp C., Kurd O.,
RA Lew H., Lin D., Mamath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE000159; AAC73632.1; -
 DR EMBL; U82598; AAB40732.1; ALT INIT.
 DR EMBL; X51662; -; NOT_ANNOTATED_CDS.
 DR Ecogene; EG12388; sfmf.
 DR InterPro: IPR000259; -
 DR Pfam; PF00419; Fimbrilal; 1.
 DR Fimbrilal; Signal; 1.
 FT SIGNAL 1 22
 FT CHAIN 23 191
 FT DISULFID 52 92
 FT SEQUENCE 191 AA; 19754 MW; EBCA2BE6FE56AF3F CRC64;

Query Match 56.3%; Score 49; DB 1; Length 191;
 Best Local Similarity 53.3%; Pred. No. 0.18;
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTALGDTAQPFSI 17
 DB 73 FNAVGTSLIPFTI 87

RESULT 5
 SFMF_ECOLI STANDARD; PRT; 171 AA.
 AC P38052; P75716; P77079;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1994 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE FIMBRIAL-LIKE PROTEIN SFMF PRECURSOR.
 GN SFMF.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_Taxid=562;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12 / MG1655;
 RC MEDLINE-9742617; PubMed-9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
 RA Lin D., Namath A., Oefner P., Roberts D., Schram S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 119-171 FROM N.A.
 RC STRAIN-K12;
 RC MEDLINE-90220507; PubMed-2183007;
 RA Muramatsu S., Mizuno T.;
 RT "Nucleotide sequence of the yna region encompassing the int gene of a
 RT cryptic prophage and the dna Y gene flanked by a curved DNA sequence
 RT of Escherichia coli K12."
 RL Mol. Gen. Genet. 220:325-328(1990).
 RN [4]
 RN IDENTIFICATION.
 RP MEDLINE-95075659; PubMed-7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a

RT bacterial genome." 22:4756-4767(1994).
 CC Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIAL PROTEINS.
 CC -1- SIMILARITY: STRONG, TO S.TYPHIMURUM FIMF.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE000159; AAC73636.1; -
 DR EMBL; U82598; AAB40732.1; ALT INIT.
 DR EMBL; X51662; -; NOT_ANNOTATED_CDS.
 DR Ecogene; EG12388; sfmf.
 DR InterPro: IPR000259; -
 DR Pfam; PF00419; Fimbrilal; 1.
 DR Fimbrilal; Signal; 1.
 FT SIGNAL 1 20
 FT CHAIN 21 171
 FT DISULFID 40 80
 FT SEQUENCE 171 AA; 18222 MW; 59230981242A1DCA CRC64;

Query Match 54.0%; Score 47; DB 1; Length 171;
 Best Local Similarity 81.8%; Pred. No. 0.35;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 7 GDTTALVPEFSI 17
 DB 65 GDTTALVPEFSI 75

RESULT 6
 HED1_HAEIN STANDARD; PRT; 216 AA.
 ID HFD1_HAEIN
 AC P45992;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE MINOR FIMBRIAL SUBUNIT HIFD PRECURSOR.
 GN HIFD.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 CC NCBI_Taxid=727;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-EAGAN / SEROTYPE B;
 RC MEDLINE-95012708; PubMed-7927773;
 RA McCrea K.W., Watson W.J., Glasner J.D., Marris C.F.;
 RT "Identification of hifd and hife in the plus gene cluster of
 RT Haemophilus influenzae type b strain Eagan."
 RL Infect. Immun. 62:4922-4928(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-AM30 (770235) / SEROTYPE B;
 RC MEDLINE-95089703; PubMed-7997179;
 RA van Ham M.S., van Alphen L., Mooi F.R., van Putten J.P.M.;
 RT "The fimbrial gene cluster of Haemophilus influenzae type b."
 RL Mol. Microbiol. 13:673-684(1994).
 CC -1- FUNCTION: MAY BE A MINOR STRUCTURAL PROTEIN REQUIRED FOR PLUS
 CC BIOGENESIS.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```
QY      1 ASFTALGDTTAQVPFSL 17
        | : | | | | | |
Db      90 ANLSLAGQTAAPVPFSL 106
```

RESULT	7	
HFD2_HAEIN		
ID	HFD2_HAEIN	STANDARD;
MAC	P45993;	PRT; 205 AA

DPT 01-NOV-1995 (Rel. 32, Created)
 DPT 01-NOV-1995 (Rel. 32, last sequence update)
 DPT 01-NOV-1995 (Rel. 32, last annotation update)
 MINOR FIBRILLAL SUBUNIT HFD PRECURSOR.
 HFD.
 Haemophilus influenzae.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellales.
 Haemophilus
 NCBI_TaxID=727;
 [1]

1P SEQUENCE FROM N.A.
 2C STRAIN-86-0295 / LKP SEROTYPE 1;
 3A Green B.A., Olmsted S.B.;
 4A Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 5C -1- FUNCTION: MAY BE A MINOR STRUCTURAL PROTEIN REQUIRED FOR PILOSUS
 6C BIOGENESIS.
 7C
 8C -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 9C ANCHOR (PROBABLE).
 10C

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

SEQUENCE	205 AA:	21663 MW:	0DA047DAFABCDDB5 CRC64:
INTERPRO: IPR000259;	-	-	-
InterPro: IPR000437;	-	-	-
Pfam: PF00419; Fimbrial; 1.			
PROSITE: PS00013; PROKAR_LIPOPROTEIN; FALSE_NEG.			
Fimbrlia; Signal; Lipoprotein.			
SIGNAL	1	19	PROBABLE.
CHAIN	20	205	MINOR FIMBRIAL SUBUNIT HI.D.
LIPID	20	20	N-ACYL DIGLYCERIDE (PROBABLE).

Query Match	48.3%;	Score 42;	DB 1;	Length 205;
Best Local Similarity	52.9%;	Pred. No. 3.2;		
Matches	9;	Conservative	2;	Mismatches 6; Totals 0.

QY	1	ASFTAIGDTTAQVPFSI	17
		: :	
Db	96	ANLSHAGQTALPVPFSI	112

RESULT	8	
LPEA_SALTY		
ID	LPEA_SALTY	STANDARD;
AC	P43660;	PRT; 178 AA

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE LONG POLAR FIMBRIA PROTEIN A PRECURSOR.
GN LPFA.

Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
[11]
NN

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=95238281; PubMed=7721701;
RA Bauemler A.J.; Heffron F.

RT fimbrial operon of *Salmonella typhimurium*.⁸; J. Bacteriol. 177:2087-2097(1995).

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[illegible]

Query Match	47.1%;	Score 41;	DB 1;	Length 178;
Best Local Similarity	53.3%;	Pred. No. 4.2;		
Matches	8;	Conservative	2.	Mismatches

```

3  FTAIGDTPAQVPFSI 17
   | | | | : : | |
64 FKAIGDKSSSKPFQI 78

```

(S)LT	9
V_BPP2	
VVP_BPP2	
P313A0;	STANDARD; PRT; 211 AA.
01-JUL-1993 (Rel. 26, Created)	
01-OCT-1996 (Rel. 34, Last sequence update)	
30-MAY-2000 (Rel. 39, Last annotation update)	
BASILEPATE ASSEMBLY PROTEIN V (GPV).	
V.	

Bacteriophage P2.
Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae; P2-like viruses.

```
NCBI_TaxID=10679;
```

SEQUENCE FROM N.A.

MEDLINE=96036485;

Haggaard-Ljungquist E., Jacobsen E., Rishovd S., Six E.W., Nilsson O
Sunshine M.C., Tiedeman

us-09-646-043-1.rsp

Thu Jun 21 08:38:34 2001

Calendar R.; genes involved in baseplate assembly.";
 RA "Bacteriophage P2: genes involved in baseplate assembly.";
 RT Virology 213:109-121(1995).
 RL (2)
 RN SEQUENCE OF 1-72 FROM N.A.
 RP MEDLINE-85160858; PubMed-3981640;
 RX Christie G.E.; Calendar R.;
 RA "Bacteriophage P2 late promoters. II. Comparison of the four late
 RT promoter sequences.";
 RL J. Mol. Biol. 181:373-382(1985).
 RN REVISIONS, SEQUENCE FROM N.A.
 RP Christie G.E.;
 RA Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RL (4)
 RN SEQUENCE OF 1-24 FROM N.A.
 RP MEDLINE-94233699; PubMed-8178426;
 RX Linderoth N.A.; Julien B.; Flick K.E.; Calendar R.; Christie G.E.;
 RA "Molecular cloning and characterization of bacteriophage P2 genes R
 RT and S involved in tail completion.";
 RL Virology 200:347-359(1994).
 CC -1- FUNCTION: FORMS THE SMALL SPIKE AT THE TIP OF THE TAIL.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF063097; AAD03282.1;
 CC DR SEQUENCE 211 AA; 22244 MW; B1CA7E47B19B91FB CRC64;
 SO
 Query Match 47.1%; Score 41; DB 1; Length 211;
 Best Local Similarity 53.6%; Pred. No. 5;
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 OY 1 ASPTAGDTTAAVPSFV 18
 DB 127 ASVTAGSSTRTVPPVWV 144
 RESULT 10
 FIML_ECOLI STANDARD; PRT; 215 AA.
 ID FIML_ECOLI
 AC P39264;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE FIMBRIN-LIKE PROTEIN FIM1.
 GN FIM1.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCB1_TaxID=562;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12 / MG1655;
 RC MEDLINE-95334362; PubMed-7610040;
 RX Burdand V.D.; Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RN Nucleic Acids Res. 23:2105-2119(1995).
 RL [2]
 RN PRELIMINARY SEQUENCE OF 1-116 FROM N.A.
 RP MEDLINE-84285425; PubMed-6147250;
 RX Klemm P.;
 RA "The fima gene encoding the type-1 fimbrial subunit of Escherichia
 RT coli. Nucleotide sequence and primary structure of the protein.";
 RL Eur. J. Biochem. 143:395-399(1984).

 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIAL PROTEINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U14003; AAA97211.1;
 CC DR EMBL: AE005002; AAC77211.1;
 CC DR EMBL: X00981; NOT_ANNOTATED_CDS.
 CC DR EcoGene: EG11974; FIM1.
 CC DR InterPro: IPR000259;
 CC DR Pfam: PF00419; Fimbrin1.1.
 CC DR Fimbrin.
 CC DR DISULFID
 CC SEQUENCE 79 119 BY SIMILARITY.
 CC 215 AA; 23583 MW; 601F727DE4864B3 CRC64;
 SO
 Query Match 47.1%; Score 41; DB 1; Length 215;
 Best Local Similarity 53.3%; Pred. No. 5; Indels 4; Gaps 0;
 Matches 8; Conservative 3; Mismatches 4;
 OY 3 FTAIGDTTAAVPSFV 17
 DB 100 FHAVGESAPVPEV 114
 RESULT 11
 VIL_HPV31 STANDARD; PRT; 504 AA.
 ID VIL_HPV31
 AC P17388;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MAJOR CAPSID PROTEIN L1.
 GN L1.
 OS Human papillomavirus type 31.
 CC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 CC NCB1_TaxID=10585;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-89299478; PubMed-2545036;
 RX Goldspink M.D.; Dislyestre D.; Temple G.F.; Lorincz A.T.;
 RA "Nucleotide sequence of human papillomavirus type 31: a cervical
 RT neoplasia-associated virus.";
 RL Virology 171:306-311(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J04353; AAA6956.1;
 CC DR PIR: G32444; P1WL31.
 CC DR InterPro: IPR002210;
 CC DR Pfam: PF00500; Late_protein_L1.1.
 CC DR PRINTS: PR000865; HPVCAPSIDL1.
 CC DR Coat protein; Late protein.
 CC SEQUENCE 504 AA; 56352 MW; B45A306A6B3AB9D2 CRC64;
 SO
 Query Match 47.1%; Score 41; DB 1; Length 504;
 Best Local Similarity 53.3%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 3 FTAGDTTAAQVPSI 17
 Db 211 FTALODTKSNVPLDI 225

RESULT 12

FIMF_ECOLI
 ID FIMF_ECOLI STANDARD; PRT; 176 AA.
 AC P08189;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE FIMF PROTEIN PRECURSOR.
 GN FIMF
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 ON NCBI_Taxid=562;
 RX MEDLINE=88038337; PubMed=2890081;
 RT Klemm P., Christiansen G.;
 RT "Three fim genes required for the regulation of length and mediation
 of adhesion of Escherichia coli type 1 fimbriae.";
 RL Mol. Gen. Genet. 208:439-445(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]
 RP SEQUENCE OF 1-3 FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE=90220509; PubMed=1970114;
 RA Klemm P., Christiansen G.;
 RT "The fim gene required for cell surface localization of Escherichia
 coli type 1 fimbriae.";
 RL Mol. Gen. Genet. 220:334-338(1990).
 RP ADHESION INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
 OF FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
 CC -1- SIMILARITY: INVOLVED IN THE INTEGRATION OF FIMH IN THE FIMBRIAE.
 CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 DR EMBL; X05672; CAA29154.1;
 DR EMBL; U14003; AAA97214.1;
 DR EMBL; AE000502; AAC7274.1;
 DR EMBL; X51655; CAA35969.1;
 DR PIR; S07321; S07321.
 DR Ecogene; EG10313; fImf.
 DR InterPro; IPR000259; fImf.
 DR Pfam; PF00419; fImbrial; 1.
 DR Fimbrin; Signal; 1.
 FT CHAIN 1 20
 FT DISULFID 1 176
 FT SITE 38 78
 FT 175 175
 FT POTENTIAL.
 FT FIMF PROTEIN.
 FT PROBABLE.
 FT REQUIRED FOR STABILITY AND TRANSPORT (BY
 FT SIMILARITY)
 FT P -> S (IN REF. 1).
 FT S -> L (IN REF. 1).
 FT A -> V (IN REF. 1).
 FT AVKVGFTGV -> RKKGVYWR (IN REF. 1).
 FT CONFLICT 70 70
 FT CONFLICT 76 76
 FT CONFLICT 81 81
 FT CONFLICT 84 92

FT CONFLICT 107 107
 SQ SEQUENCE 176 AA; 18715 MW; 38692EPE6A40121P CRC64;

Query Match
 Best Local Similarity 46.0%; Score 40; DB 1; Length 176;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 FTAGDTTAAQVPSI 18
 Db 59 FNNIGATVPEVPRIL 74

RESULT 13

HAGA_PORGI
 ID HAGA_PORGI STANDARD; PRT; 2628 AA.
 AC Q51845;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE HEMAGGLUTININ A PRECURSOR.
 GN HAGA
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; CF8 group; Bacteroidaceae; Porphyromonas.
 ON NCBI_Taxid=837;
 RX MEDLINE=97047672; PubMed=8926061;
 RT Han N., Whitlock T., Prognulski-Fox A.;
 RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
 infect. Immun. 64:4000-4007(1996).
 RL -1- FUNCTION: AGGLUTININATES ERYTHROCYTES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C25.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 DR EMBL; U41807; AAB17128.1;
 DR InterPro; IPR001769; fImf.
 DR Pfam; PF01364; Peptidase_C25; 5.
 DR Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
 FT CHAIN 1 24
 FT DOMAIN 25 2628
 FT DOMAIN 25 539
 FT DOMAIN 540 995
 FT DOMAIN 996 1451
 FT DOMAIN 1452 1907
 FT DOMAIN 2074 2628
 FT SEQUENCE 2628 AA; 283324 MW; 61C4DE32340C99DA CRC64;
 Query Match
 Best Local Similarity 46.0%; Score 40; DB 1; Length 2628;
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 Oy 2 SFTAGDTTAAQVPSI 17
 Db 2554 NITSADVTAQKPYTL 2569
 RESULT 14
 EXON_HSV6U STANDARD; PRT; 488 AA.
 AC P24447;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)

Thu Jun 21 08:38:34 2001

us-09-646-043-1.rsp

Page 7

01-OCT-2000 (rel. 40, last annotation update)
DE ALKALINE EXONUCLEASE (EC 3.1.11.-).
GN U70 OR 16R.
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=10370;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90080132; PubMed=2152817;
RA Lawrence G.L., Chee M., Craxton M.A., Compels U.A., Honess R.W.,
RT J. Virol. 64:287-299(1990).
RL "Human herpesvirus 6 is closely related to human cytomegalovirus."
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95266321; PubMed=7747482;
RA Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RT Martin M.E., Efsthliou S., Craxton M., Macaulay H.A.;
RL "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RT Virology 209:29-51(1995).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES ALKALINE EXONUCLEASE
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X83413; CAA58362.1; -
CC DR EMBL: M68963; AAB5578.1; -
CC DR PIR: F36769; COBHS.
CC DR InterPro: IPR001616; -
CC DR Pfam: PF01771; Herpes_alik_exo; 1.
CC DR PRINTS: PR00924; ALKEXNUCLASE.
CC DR HydroLase: Nuclease; Exonuclease.
CC KW SEQUENCE 488 AA; 56644 MW; 0F38A10597366A5B CRC64;
SQ
Query Match 44.8%; Score 39; DB 1; Length 488;
Best Local Similarity 64.3%; Pred. No. 28;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 5 AIGDTAOPFSIV 18
|| : |||||
DB 29 AIREKIKOPFSIV 42
RESULT 15
EXON_HSV62 STANDARD: PRT; 488 AA.
ID EXON_HSV62
AC P52448;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALKALINE EXONUCLEASE (EC 3.1.11.-).
GN U70 OR CH3R.
OS Human herpesvirus (type 6 / strain 229) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=16351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8634027; PubMed=8634027;
RA Lindquester G.J., Inoue N., Allen R.D., Castelli J.W.,
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,
RA Frenkel N., Pellett P.E.;
RT "Restriction endonuclease mapping and molecular cloning of the human
herpesvirus 6 variant B strain 229 genome."
RT

RL Arch. Virol. 141:367-379(1996).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES ALKALINE EXONUCLEASE
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF157706; AAB06353.1; -
CC DR InterPro: IPR001616; Herpes_alik_exo; 1.
CC DR Pfam: PF01771; Herpes_alik_exo; 1.
CC DR PRINTS: PR00924; ALKEXNUCLASE.
CC DR HydroLase: Nuclease; Exonuclease.
CC KW SEQUENCE 488 AA; 56687 MW; AE2872028D4B3D90 CRC64;
SQ
Query Match 44.8%; Score 39; DB 1; Length 488;
Best Local Similarity 64.3%; Pred. No. 28;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 5 AIGDTAOPFSIV 18
|| : |||||
DB 29 AIREKIKOPFSIV 42
Search completed: June 20, 2001, 13:32:54
Job time: 92 sec

Thu Jun 21 08:38:34 2001

us-09-646-043-1.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 13.45 Seconds
(without alignments)
101.944 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASFTAGDTPAQVPFSIV 18

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	184	2	S20682
2	82	94.3	185	2	B28393
3	49	56.3	191	2	A64785
4	49	56.3	191	2	C85553
5	47	54.0	171	2	B64785
6	44.5	51.1	591	2	S65588
7	44	50.6	171	2	G85553
8	44	50.6	931	2	B85062
9	43	49.4	216	2	S54430
10	43	49.4	659	2	T33557
11	43	49.4	1248	2	B96827
12	42	48.3	194	2	G85976
13	42	48.3	144	2	T41416
14	42	48.3	1146	2	E70204
15	41	47.1	177	2	A56271
16	41	47.1	199	2	A61725
17	41	47.1	215	2	S56540
18	41	47.1	215	2	E86129
19	41	47.1	504	1	PIWL31
20	41	47.1	628	2	T84239
21	41	47.1	1298	2	T77523
22	40	46.0	176	2	S56543
23	40	46.0	176	2	H86129
24	40	46.0	1419	2	T32970
25	40	46.0	1732	2	T30836
26	40	46.0	2628	2	T28651
27	39	44.8	270	2	T33039
28	39	44.8	409	2	S77262
29	39	44.8	468	2	B70932

30	39	44.8	488	1	O0B8H5	alkaline exonuclease
31	39	44.8	488	2	T44030	alkaline exonuclease
32	39	44.8	488	2	T44215	alkaline exonuclease
33	39	44.8	615	2	T39758	proline-serine ric
34	39	44.8	725	2	E86790	hypothetical prote
35	38	43.7	85	2	T17696	hypothetical prote
36	38	43.7	170	2	I39892	hypothetical prote
37	38	43.7	175	2	S06193	flag protein precu
38	38	43.7	175	2	I59446	flag protein - Esc
39	38	43.7	178	2	C86029	probable major flm
40	38	43.7	202	2	C71306	hypothetical prote
41	38	43.7	218	2	F75404	ABC transporter, p
42	38	43.7	259	2	T23782	hypothetical prote
43	38	43.7	347	2	S43771	phosphatidylcholin
44	38	43.7	458	2	T27217	hypothetical prote
45	38	43.7	460	2	A84281	hypothetical prote

ALIGNMENTS

RESULT 1
S20682
type 1 fimbrial protein flmA - Salmonella typhi
C:Species: Salmonella typhi
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C:Accession: S20682
R:Rosso, G.M.; Muscas, P.; Chiesurin, A.; Satta, G.
submitted to the EMBL Data Library, March 1992
A:Reference number: S20682
A:Accession: S20682
A:Molecule type: DNA
A:Residues: 1-184 <ROS>
A:Cross-references: EMBL:X65168; NID:9395382; PIDN:CAA46286.1; PID:947667
A:Experimental source: strain Sty4
C:Genetics:
A:Gene: flmA
C:Superfamily: type 1 fimbrial protein
C:Keywords: fimbria

Query Match 100.0%; Score 87; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 8.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ASFTAGDTPAQVPFSIV 18
DB 65 ASFTAGDTPAQVPFSIV 82
RESULT 2
B28393
type 1 fimbrial protein precursor - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 15-Nov-1996
C:Accession: B28393; A05121
R:Purcell, B.K.; Pruckler, J.; Clegg, S.
J. Bacteriol. 169, 5831-5834, 1987
A:Title: Nucleotide sequences of the genes encoding type 1 fimbrial subunits of Klebs
A:Reference number: A91858; WUID:88058806
A:Accession: B28393
A:Molecule type: DNA
A:Residues: 1-185 <PDR>
A:Note: The authors translated the codon GCC for residue 18 as Gly, GTC for residue 3
R:Maalen, K.; Sletten, K.; Froholm, L.O.; Valsanen, V.; Korhonen, T.K.
FEMS Microbiol. Lett. 16, 149-151, 1983
A:Reference number: A05121
A:Accession: A05121
A:Molecule type: protein
A:Residues: 23-30, 'S', '32-45', 'X', '47-50' <MAA>
C:Superfamily: type 1 fimbrial protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-185/Product: type 1 fimbrial protein #status predicted <MAT>

Query Match 94.3%; Score 82; DB 2; Length 185;
 Best Local Similarity 94.4%; Pred. No. 6.5e-07;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASPTAIGDTTAQVPFSIV 18
 |||||:|||||
 DB 65 ASPTAIGDTTAQVPFSIV 82

RESULT 3

A64785
 type 1 fimbrial protein sfm precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 26-Aug-1999

C:Accession: A64785
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC

A:Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617

A:Accession: A64785

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-191 <BLAT>

A:Cross-references: GB:AE000159; GB:U00096; NID:g1786739; PIDN:AACT3632.1; PID:g1786742;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
 A:Gene: sfm

C:Superfamily: type 1 fimbrial protein

C:Keywords: fimbria; mannose-resistant hemagglutination

OY 3 FTAIGDTTAQVPFSI 17
 |||||:|||||
 DB 73 FNAVGNTSALIPFTI 87

RESULT 4

C85553
 probable fimbrial-like protein sfm [imported] - Escherichia coli (strain 0157:H7)

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: C85553
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: C85553

A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-191 <STO>

A:Cross-references: GB:AE005174; NID:g12513430; PIDN:AAG54887.1; GSPDB:GNO0145; UWGP:206

A:Experimental source: strain 0157:H7, substrain EDL933

C:Genetics:
 A:Gene: sfm

C:Superfamily: type 1 fimbrial protein

Query Match 56.3%; Score 49; DB 2; Length 191;
 Best Local Similarity 53.3%; Pred. No. 0.4;
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 5

E64785
 fimbrial protein homolog sfm - Escherichia coli

C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 18-Aug-2000

C:Accession: E64785

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A:Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617

A:Accession: E64785

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-171 <BLAT>

A:Cross-references: GB:AE000159; GB:U00096; NID:g1786739; PIDN:AACT3636.1; PID:g17867

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
 A:Gene: sfm; ybcg

C:Superfamily: type 1 fimbrial protein

Query Match 54.0%; Score 47; DB 2; Length 171;
 Best Local Similarity 81.8%; Pred. No. 0.79;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 7 GDTTAQVPFSI 17
 |||||:|||||
 DB 65 GDTTALVPFSL 75

RESULT 6

S65588
 ABC-transporter Strw - Streptomyces glaucescens

C:Species: Streptomyces glaucescens
 C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 02-Feb-2001

C:Accession: S65588; S57562

R:Beyer, S.; Distler, J.; Piepersberg, W.
 Mol. Gen. Genet. 250, 775-784, 1996

A:Title: The str gene cluster for the biosynthesis of 5'-hydroxystreptomycin in Strept

A:Reference number: S65585; MUID:96204519

A:Accession: S65588

A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-591 <BE>

A:Cross-references: EMBL:X89010; NID:g887633; PIDN:CAAG1417.1; PID:g887637

C:Genetics:
 A:Gene: strw

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

C:Keywords: ATP; nucleotide binding; P-loop
 F:358-551/Domain: ATP-binding cassette homology <ABC>

F:375-382/Region: nucleotide-binding motif A (P-loop)

Query Match 51.1%; Score 44.5; DB 2; Length 591;
 Best Local Similarity 66.7%; Pred. No. 8.8;
 Matches 10; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

OY 1 ASPTAIGDTTAQVP 14
 |||||:|||||
 DB 200 ASPTAIGDTTAQVP 214

RESULT 7

G85553
 probable fimbrial protein sfm [imported] - Escherichia coli (strain 0157:H7)

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: G85553

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: G85553
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-171 <STO>
 A:Cross-references: GB:AE005174; NID:912513434; PIDN:AG54891.1; GSPDB:GN00145; UWGP:206
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: stx2f
 C:Superfamily: type 1 fimbrial protein

Query Match
 Best Local Similarity 50.6%; Score 44; DB 2; Length 171;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 7 GDTTAAQVPSI 17
 |||||:
 DB 65 GDTTALLPFSL 75

RESULT 8
 B85062
 probable WD-repeat membrane protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
 C:Accession: B85062
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 A:Reference number: AB5001; MUID:20083488
 A:Accession: B85062
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-931 <STO>
 A:Cross-references: GB:NC_001268; NID:97267253; PIDN:CAB81036.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4G04940
 A:Map position: 4

Query Match
 Best Local Similarity 50.6%; Score 44; DB 2; Length 931;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 3 FRAIGDTTAAQVPSI 17
 |||||:
 DB 7 FRAIGTITVPSV 21

RESULT 9
 S54430
 fimbrial protein hlfD precursor - Haemophilus influenzae (strain AM30)
 C:Species: Haemophilus influenzae
 C:Date: 06-Sep-1996 #sequence_revision 13-Mar-1997 #text_change 24-Nov-1999
 C:Accession: S54430
 R:Van Ham, S.M.; Van Alphen, L.; Mool, F.R.; van Putten, J.P.M.
 Mol. Microbiol. 13, 673-684, 1994
 A:Title: The fimbriae gene cluster of Haemophilus influenzae type b.
 A:Reference number: S54428; MUID:95089703
 A:Accession: S54430
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-216 <VAN>
 A:Cross-references: EMBL:Z33502; NID:9535165; PIDN:CAA83903.1; PID:9535168
 C:Genetics:
 A:Gene: hlfD
 C:Superfamily: type 1 fimbrial protein

Query Match 49.4%; Score 43; DB 2; Length 216;

Best Local Similarity 52.9%; Pred. No. 5.2;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 1 ASFTAGDTTAAQVPSI 17
 |||||:
 DB 90 ANLSLAGTAAQVPSI 106

RESULT 10
 T33557
 hypothetical protein W04C9.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33557
 R:Fulton, R.; Rohlfing, T.; Morris, M.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of C. elegans cosmid W04C9.
 A:Reference number: Z21368
 A:Accession: T33557
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-659 <FUL>
 A:Cross-references: EMBL:AR098999; PIDN:AC68729.1; GSPDB:GN00019; CESP:W04C9.6
 A:Experimental source: strain Bristol N2; clone W04C9
 A:Gene: CESP:W04C9.6
 A:Map position: 1
 A:Introns: 11/3; 48/1; 113/1; 240/3; 320/2; 361/3; 393/3; 417/2; 477/2

Query Match 49.4%; Score 43; DB 2; Length 659;
 Best Local Similarity 56.2%; Pred. No. 18;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 2 SFTAGDTTAAQVPSI 17
 |||||:
 DB 60 SCVAIGSLAATPSI 75

RESULT 11
 B96827
 hypothetical protein TBK14.1 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: B96827
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Malt, R.; Marzia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719
 A:Accession: B96827
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1248 <STO>
 A:Cross-references: GB:AE005173; NID:94835752; PIDN:AAD30219.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: TBK14.1
 A:Map position: 1

Query Match 49.4%; Score 43; DB 2; Length 1248;
 Best Local Similarity 42.9%; Pred. No. 38;
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ASFTAGDTTAAQV 14
 |||||:
 DB 464 SNYSAGTSSIP 477

RESULT 12

probable fimbrial-like protein yrah [imported] - Escherichia coli (strain O157:H7)
 G55976
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: G85976
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dialantza, E.; Potamoulsis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-194 <STO>
 A:Cross-references: GB:AE005174; NID:g12517744; PIDN:AA658275.1; GSPDB:GN00145; UMGF:244
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: yrah

Query Match 48.3%; Score 42; DB 2; Length 194;
 Best Local Similarity 40.0%; Pred. No. 6.9;
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFSI 17
 I : I I : : I I : I
 DB 59 FRSVODRSPKIPFTI 73

RESULT 13

T41416
 probable tyrosine-tRNA ligase precursor, mitochondrial - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T41416
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21954
 A:Accession: T41416
 A:Status: preliminary; translated from GB/EMBL/DBDJB
 A:Molecule type: DNA
 A:Residues: 1-445 <WOO>
 A:Cross-references: EMBL:AL031798; PIDN:CAA21185.1; GSPDB:GN00068; SPDB:SPCC576.06c
 A:Experimental source: strain 972h-; cosmid c576
 C:Genetics:
 A:Gene: SPDB:SPCC576.06c
 A:Map position: 3
 A:Genome: nuclear
 A:Introns: 22/2; 268/3
 C:Superfamily: tyrosine-tRNA ligase
 C:Keywords: mitochondrion

Query Match 48.3%; Score 42; DB 2; Length 445;
 Best Local Similarity 72.7%; Pred. No. 18;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQV 13
 I I I I I I I I I I
 DB 65 FTVIGDATAQL 75

RESULT 14

E70204
 hypothetical protein B80838 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: E70204
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Iathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:36065943
 A:Accession: E70204
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1146 <KLE>
 A:Cross-references: GB:AE001182; GB:AE000783; NID:g2688780; PIDN:AA67188.1; PID:g268
 A:Experimental source: strain B31

Query Match 48.3%; Score 42; DB 2; Length 1146;
 Best Local Similarity 53.3%; Pred. No. 51;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFSI 17
 I : I I I I I I I I I I
 DB 948 FSKIGTTTKLVPSL 962

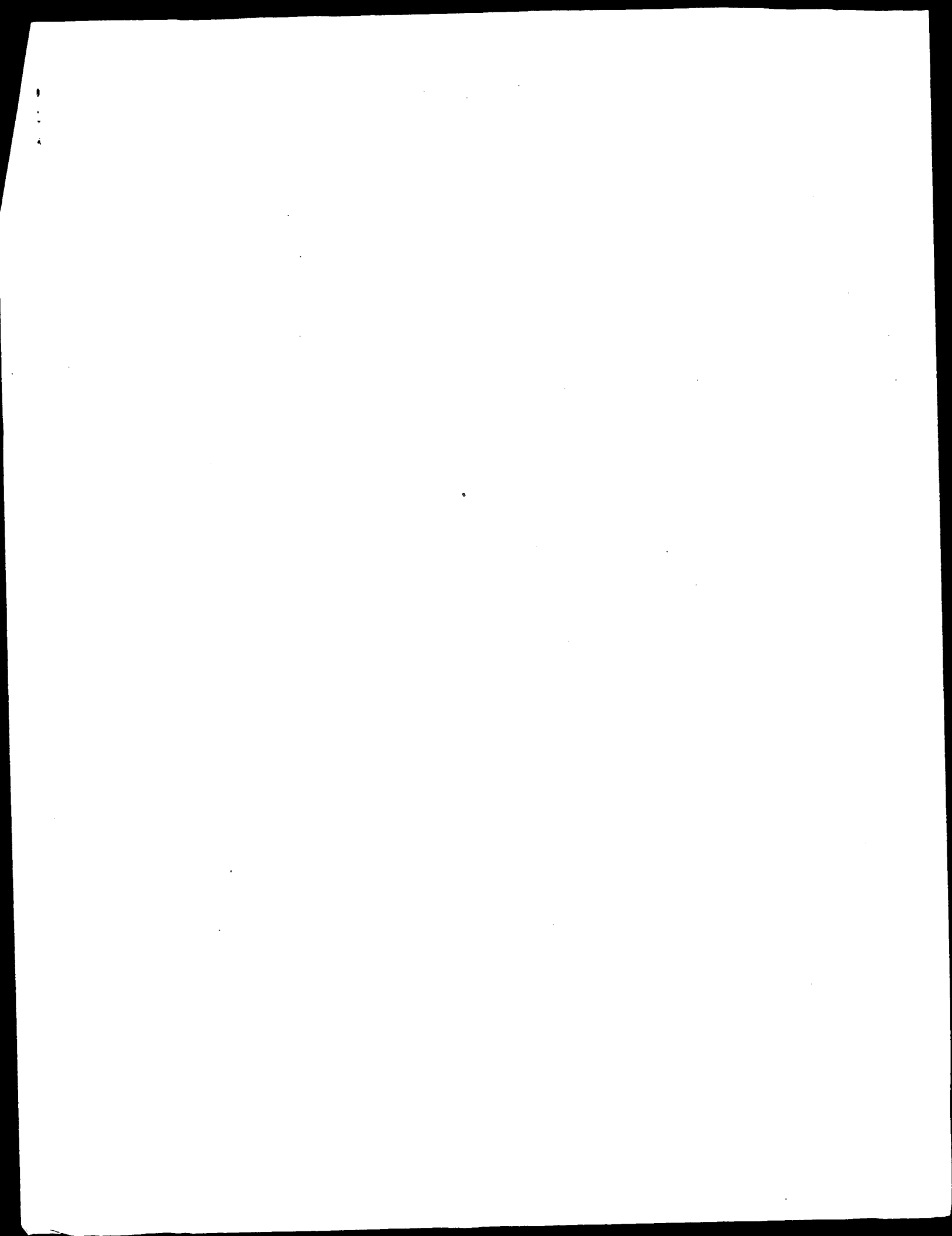
RESULT 15

A56271
 long polar fimbrial major protein precursor - Salmonella typhimurium
 C:Species: Salmonella typhimurium
 C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 07-May-1999
 C:Accession: A56271
 R:Beaumont, A.J.; Heffron, F.
 J. Bacteriol. 177, 2087-2097, 1995
 A:Title: Identification and sequence analysis of lpfABCD, a putative fimbrial operon
 A:Reference number: A56271; MUID:95238281
 A:Accession: A56271
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-177 <BAE>
 A:Cross-references: GB:U18559
 C:Genetics:
 A:Gene: lpfA
 C:Superfamily: type 1 fimbrial protein

Query Match 47.1%; Score 41; DB 2; Length 177;
 Best Local Similarity 53.3%; Pred. No. 9.2;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFSI 17
 I I I I I I I I I I
 DB 64 FKAIGDKSSSKPFOI 78

Search completed: June 20, 2001, 13:32:19
 Job time: 57 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 19.45 seconds
(without alignments)
122.442 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87
Sequence: 1 ASFTAIGDTAQPFSIV 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_16:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.prodent:*
12: sp.unclassified:*
13: sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	160	2 Q9X3T3	Q9X3T3 salmonella
2	84	96.6	161	2 Q9X3U2	Q9X3U2 salmonella
3	84	96.6	161	2 Q9WM13	Q9WM13 salmonella
4	82	94.3	161	2 Q9X3T5	Q9X3T5 salmonella
5	82	94.3	185	2 Q53483	Q53483 salmonella
6	79	90.8	161	2 Q9WM63	Q9WM63 salmonella
7	76	87.4	161	2 Q9X3U1	Q9X3U1 salmonella
8	76	87.4	161	2 Q9WVW8	Q9WVW8 salmonella
9	76	87.4	161	2 Q9S644	Q9S644 salmonella
10	72	82.8	161	2 Q9X3T9	Q9X3T9 salmonella
11	72	82.8	161	2 Q9X3T7	Q9X3T7 salmonella
12	72	82.8	161	2 Q9S643	Q9S643 salmonella
13	52	59.8	186	2 P72209	P72209 protease mir
14	46	52.9	510	10 Q9FX9	Q9FX9 glycine mtr
15	46	52.9	204	2 Q9LRM3	Q9LRM3 arabidopsis
16	45	51.7	216	2 Q9K3U2	Q9K3U2 streptomyc
17	45	51.1	2	2 Q54204	Q54204 streptomyc
18	44.5	50.6	931	10 Q9ZPH3	Q9ZPH3 arabidopsis
19	44				

20	43	49.4	221	2 P94814	P94814 haemophilus
21	43	49.4	659	5 Q9TZE1	Q9TZE1 caenorhabdi
22	43	49.4	1248	10 Q9SAJ2	Q9SAJ2 arabidopsis
23	42	48.3	292	2 Q85245	Q85245 thermotoga
24	42	48.3	359	5 Q9GUB4	Q9GUB4 giardia lam
25	42	48.3	445	5 Q74890	Q74890 schistosacch
26	42	48.3	1146	2 Q51778	Q51778 borrelia bu
27	41	47.1	199	2 Q9PL66	Q9PL66 chlamydia bu
28	41	47.1	210	2 Q87338	Q87338 chlamydia m
29	41	47.1	628	14 Q81017	Q81017 human papill
30	41	47.1	504	1 Q9HRN6	Q9HRN6 halobacteri
31	41	47.1	1298	10 Q9LZU7	Q9LZU7 arabidopsis
32	40	46.0	176	2 Q9F626	Q9F626 escherichia
33	40	46.0	176	2 Q9F626	Q9F626 escherichia
34	40	46.0	205	2 Q87470	Q87470 haemophilus
35	40	46.0	1097	2 P72196	P72196 porphyromon
36	40	46.0	1223	2 Q9ZNB5	Q9ZNB5 porphyromon
37	40	46.0	1419	5 Q45092	Q45092 caenorhabdi
38	40	46.0	1732	2 Q51817	Q51817 porphyromon
39	40	46.0	1732	2 Q52050	Q52050 porphyromon
40	40	46.0	1732	2 Q07442	Q07442 porphyromon
41	39	44.8	201	2 Q9K3U2	Q9K3U2 streptomyc
42	39	44.8	205	2 Q86217	Q86217 haemophilus
43	39	44.8	358	10 Q9ZP40	Q9ZP40 plasmid sativ
44	39	44.8	409	2 P73556	P73556 synechocyst
45	39	44.8	411	5 Q9VM76	Q9VM76 drosophila

ALIGNMENTS

RESULT 1					
Q9X3T3		PRELIMINARY:	PRT:	160 AA.	
AC Q9X3T3:					
DT 01-NOV-1999 (TREMBLrel. 12, Created)					
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)					
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)					
DE MAJOR PILIN PROTEIN FILA (FRAGMENT).					
GN FILA.					
OS Salmonella enterica subsp. enterica serovar Typhl.					
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;					
OC Salmonella.					
OX NCBI_TaxID=90370;					
RN [1]					
RP SEQUENCE FROM N.A					
RC STRAIN=K83333;					
RX MEDLINE=99138752; PubMed=9973358;					
RA Boyd E.F., Hartl D.L.;					
RT "Analysis of the type 1 pilin gene cluster flm in Salmonella: its					
RL distinct evolutionary histories in the 5' and 3' regions.";					
DR J. Bacteriol. 181:1301-1308(1999)					
DR EMBL: AF063899; AM239211					
DR InterPro: IPR000259; .					
DR Pfam: PF00419; Pimbr1al; 1.					
FT NON_TER					
SO SEQUENCE					
Query Match		100.0%;	Score 87;	DB 2;	Length 160;
Best Local Similarity		100.0%;	Pred. No. 3.6e-07;		
Matches 18;		Conservative 0;	Mismatches 0;	Gaps 0;	
QY 1 ASFTAIGDTAQPFSIV 18					
DB 41 ASFTAIGDTAQPFSIV 58					
RESULT 2					
Q9X3U2		PRELIMINARY:	PRT:	161 AA.	
AC Q9X3U2:					
DT 01-NOV-1999 (TREMBLrel. 12, Created)					

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. salamae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=59202;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-RKS2985;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster in Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083911; AAD23945.1; -.
 DR InterPro; IPR000259; -.
 DR Pfam; PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16441 MW; BE3DA852B6144AA2 CRC64;

Query Match 96.6%; Score 84; DB 2; Length 161;
 Best Local Similarity 94.4%; Pred. No. 1.2e-06;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQPFSIV 18
 |||||
 DB 41 ASFTAGDTTAQPFSIV 58

RESULT 3
 O9WM13 PRELIMINARY; PRT; 161 AA.
 AC 09WM13;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. diarizonae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=59204;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-RKS2979; RKS2978;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster in Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083902; AAD23927.1; -.
 DR EMBL; AF083901; AAD23925.1; -.
 DR InterPro; IPR000259; -.
 DR Pfam; PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16515 MW; 4A7ELEF9ACA5E1B8 CRC64;

Query Match 96.6%; Score 84; DB 2; Length 161;
 Best Local Similarity 94.4%; Pred. No. 1.2e-06;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQPFSIV 18
 |||||
 DB 41 ASFTAGDTTAQPFSIV 58

RESULT 4
 O9X3T5 PRELIMINARY; PRT; 161 AA.
 ID O9X3T5

AC O9X3T5;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. enterica serovar Typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=90371;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-RKS4194;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster in Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083900; AAD23923.1; -.
 DR InterPro; IPR000259; -.
 DR Pfam; PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16422 MW; 372FA51E6018939A CRC64;

Query Match 94.3%; Score 82; DB 2; Length 161;
 Best Local Similarity 94.4%; Pred. No. 2.5e-06;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQPFSIV 18
 |||||
 DB 41 ASFTAGDTTAQPFSIV 58

RESULT 5
 O33483 PRELIMINARY; PRT; 185 AA.
 AC O33483;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE FIMA.
 GN FIMA.
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=592;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC MEDLINE=95174773; PubMed=7870071;
 RA Doran J.L., Collinson S.K., Kay C.M., Baner P.A., Burian J.,
 RA Munro C.K., Lee S.H., Somers J.M., Todd E.C., Kay W.W.;
 RT "Fima and tctc based DNA diagnostics for Salmonella."
 RL Mol. Cell. Probes 8:291-310(1994).
 DR EMBL; S76043; AAB33536.1; -.
 DR InterPro; IPR000259; -.
 DR Pfam; PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 185 AA; 18870 MW; 1466F94DE872FB87 CRC64;

Query Match 94.3%; Score 82; DB 2; Length 185;
 Best Local Similarity 94.4%; Pred. No. 2.9e-06;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQPFSIV 18
 |||||
 DB 65 ASFTAGDTTAQPFSIV 82

RESULT 6
 O9WM63 PRELIMINARY; PRT; 161 AA.
 ID O9WM63
 AC O9WM63;

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. indica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OC NCBI_TaxID=59207;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS2955; RKS3057;
 RX MEDLINE-99138752; PubMed-9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster flm in Salmonella: its
 RT distinct evolutionary histories in the 5' and 3' regions.";
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL: AF083912; AAD23947.1; -;
 DR EMBL: AF083905; AAD23933.1; -;
 DR InterPro: IPR000259; -;
 DR Pfam: PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16410 MW; 1E3A02E81D659004 CRC64;

Query Match 90.8%; Score 79; DB 2; Length 161;
 Best Local Similarity 88.9%; Pred. No. 8.2e-06;
 Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAIGDTTAQVPSIV 18
 |||:|||||:|||||:
 DB 41 ASFKAVGDTTAQVPTII 58

RESULT 7
 O9X3U1 PRELIMINARY; PRT; 161 AA.
 AC O9X3U1;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. VII.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OC NCBI_TaxID=59208;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS3014;
 RX MEDLINE-99138752; PubMed-9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster flm in Salmonella: its
 RT distinct evolutionary histories in the 5' and 3' regions.";
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL: AF083908; AAD23939.1; -;
 DR InterPro: IPR000259; -;
 DR Pfam: PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16498 MW; C868E19251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAIGDTTAQVPSIV 18
 |||:|||||:|||||:
 DB 41 ASFKAVGDTTAQVPTII 58

RESULT 8
 O9WV8

ID O9WV8 PRELIMINARY; PRT; 161 AA.
 AC O9WV8;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. houtenae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OC NCBI_TaxID=59205;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS3027; RKS3015;
 RX MEDLINE-99138752; PubMed-9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster flm in Salmonella: its
 RT distinct evolutionary histories in the 5' and 3' regions.";
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL: AF083910; AAD23943.1; -;
 DR EMBL: AF083909; AAD23941.1; -;
 DR InterPro: IPR000259; -;
 DR Pfam: PF00419; Fimbrin; 1.
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16498 MW; C868E19251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAIGDTTAQVPSIV 18
 |||:|||||:|||||:
 DB 41 ASFKAVGDTTAQVPTII 58

RESULT 9
 O9S644 PRELIMINARY; PRT; 161 AA.
 AC O9S644;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. VII.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OC NCBI_TaxID=59208;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS3013;
 RX MEDLINE-99138752; PubMed-9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster flm in Salmonella: its
 RT distinct evolutionary histories in the 5' and 3' regions.";
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL: AF083907; AAD23937.1; -;
 DR InterPro: IPR000259; -;
 DR InterPro: IPR01029; -;
 DR Pfam: PF00419; Fimbrin; 1.
 DR ProDom: PD000316; -;
 DR NON_TER 1
 FT SEQUENCE 161 AA; 16484 MW; CC39A48251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAIGDTTAQVPSIV 18
 |||:|||||:|||||:
 DB 41 ASFKAVGDTTAQVPTII 58

RESULT 10

09X3T9 PRELIMINARY; PRT; 161 AA.
 AC 09X3T9;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. salamae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=59202;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS2993;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083906; AAD33935.1; -
 DR InterPro; IPR000259; -
 DR Pfam; PF00419; Fimbrin; 1.
 FT NON_TER
 SQ SEQUENCE 161 AA; 16512 MW; C2FF6643E381C7F3 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 0.00013;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASFTAGDGTAAVPSIV 18
 |||||:|||||:|||||
 DB 41 ASFTVGDGTALVPTIV 58

RESULT 11

09X3T7 PRELIMINARY; PRT; 161 AA.
 AC 09X3T7;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. arizonae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=59203;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS2983;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083904; AAD33931.1; -
 DR InterPro; IPR000259; -
 DR Pfam; PF00419; Fimbrin; 1.
 FT NON_TER
 SQ SEQUENCE 161 AA; 16512 MW; C2FF6643E381C7F3 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 0.00013;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASFTAGDGTAAVPSIV 18
 |||||:|||||:|||||

DB 41 ASFTVGDGTALVPTIV 58

RESULT 12

09S643 PRELIMINARY; PRT; 161 AA.
 AC 09S643;
 DT 01-MAY-2000 (T-EMBlrel. 13, Created)
 DT 01-MAY-2000 (T-EMBlrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).
 GN FIMA.
 OS Salmonella enterica subsp. arizonae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=59203;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RKS2980;
 RX MEDLINE=99138752; PubMed=9973358;
 RA Boyd E.F., Hartl D.L.;
 RT "Analysis of the type 1 pilin gene cluster fim of Salmonella: its
 distinct evolutionary histories in the 5' and 3' regions."
 RL J. Bacteriol. 181:1301-1308(1999).
 DR EMBL; AF083903; AAD233929.1; -
 DR InterPro; IPR000259; -
 DR Pfam; PF00419; Fimbrin; 1.
 FT NON_TER
 SQ SEQUENCE 161 AA; 16500 MW; C2F9D6455531D456 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;
 Best Local Similarity 77.8%; Pred. No. 0.00013;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASFTAGDGTAAVPSIV 18
 |||||:|||||:|||||
 DB 41 ASFTVGDGTALVPTIV 58

RESULT 13

P72209 PRELIMINARY; PRT; 186 AA.
 AC P72209;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE MAJOR SUBUNIT OF TYPE 1 FIMBRIN.
 GN ATFA.
 OS Proteus mirabilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 OX NCBI_TaxID=584;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H14320;
 RX Masad G., Fulkerson J.F., Watson D.C., Mobley H.L.T.;
 RA Infect. Immun. 64:0-0(1996).
 DR EMBL; Z78335; CAB01712.1; -
 DR InterPro; IPR000259; -
 DR Pfam; PF00419; Fimbrin; 1.
 SQ SEQUENCE 186 AA; 19061 MW; A2FE208F7D526EC2 CRC64;

Query Match 59.8%; Score 52; DB 2; Length 186;
 Best Local Similarity 58.8%; Pred. No. 0.37;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 ASFTAGDGTAAVPSI 17
 |||||:|||||:|||||
 DB 71 AEFTRKKGDETRIPFSI 87

RESULT 14

O9FXY9

ID O9FXY9

PRELIMINARY;

PRT: 250 AA.

AC O9FXY9;

01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001

(TREMBLrel. 16, Last sequence update)

DE 01-MAR-2001

(TREMBLrel. 16, Last annotation update)

OS Glycine max (Soybean);

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

Fabales; Fabaceae; Papilionoideae; Glycine.

NCBI_TaxID=3847;

[1]

RP SEQUENCE FROM N.A.

STRAIN=CV_WILLIAMS 82; TISSUE=EPICOTYL;

RX MEDLINE=20165944; PubMed=10701117;

Graham M.A., Marek L.F., Lohnes D., Cregan P., Shoemaker R.C.;

RT "Expression and genome organization of resistance gene analogs in

soybean."

RL Genome 43:86-93(2000).

EMBL: AF175397; AAC01054.1; -.

DR NON_TER

1

FT SEQUENCE

250 AA; 28320 MW; A4A503ED063A415 CRC64;

Query Match

Best Local Similarity 52.9%; Score 46; DB 10; Length 250;

Matches 12; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

OY 4 TAIGDTTA-OVPSIV 18

| | | | |

DB 86 TLISDNTAVKOVPSIV 102

RESULT 15

O9LRW3

ID O9LRW3

PRELIMINARY;

PRT: 510 AA.

AC O9LRW3;

01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000

(TREMBLrel. 15, Last sequence update)

DE 01-OCT-2000

(TREMBLrel. 15, Last annotation update)

OS Arabidopsis thaliana (Mouse-ear cress);

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicaceae; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

[1]

RP SEQUENCE FROM N.A.

STRAIN=COLUMBIA;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20277480; PubMed=10819329;

Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

features of the regions of 4,504,864 bp covered by sixty pl and TAC

RT clones."

DNA Res. 7:131-135(2000).

DR EMBL: AB028610; BAB02908.1; -.

[1]

SO SEQUENCE

510 AA; 57181 MW; 559F8C997862866F CRC64;

Query Match

Best Local Similarity 52.9%; Score 46; DB 10; Length 510;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 4 TAIGDTTAOVPSI 17

| | | | |

DB 293 TVISDLTAADIPESV 306

Search completed: June 20, 2001, 13:32:42
Job time: 80 sec

Thu Jun 21 08:38:35 2001

us-09-646-043-1.rsp

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:39:48 ; Search time 45.83 seconds

(without alignments)

13.708 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASPTAIGDTTAAQVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1412450 seqs, 34903136 residues

Total number of hits satisfying chosen parameters: 1412450

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New:**

1: /cgn2_6/ptodata/1/paa/PC7_NEW_COMB.pep:**

2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:**

3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:**

4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:**

5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:**

6: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	18	US-09-646-043-1	Sequence 1, Appl1
2	37	42.5	557	US-09-134-001C-5569	Sequence 5569, Ap
3	36	41.4	224	US-09-134-001C-4010	Sequence 4010, Ap
4	36	41.4	293	US-09-134-001C-3803	Sequence 3803, Ap
5	34	39.1	503	US-09-134-001C-4214	Sequence 4214, Ap
6	33	37.9	333	US-09-454-684A-292	Sequence 292, App
7	33	37.9	333	US-09-620-412A-292	Sequence 292, App
8	33	37.9	476	US-09-134-001C-3778	Sequence 3778, Ap
9	33	37.9	901	US-09-134-001C-5351	Sequence 5351, Ap
10	32	36.8	25	US-09-724-059-562321	Sequence 562321, Ap
11	32	36.8	25	US-09-724-059-564841	Sequence 564841, Ap
12	32	36.8	25	US-09-724-059-567241	Sequence 567241, Ap
13	32	36.8	25	US-09-724-059-569761	Sequence 569761, Ap
14	32	36.8	25	US-09-724-059-574561	Sequence 574561, Ap
15	32	36.8	123	US-09-724-059-579601	Sequence 579601, Ap
16	32	36.8	177	US-09-134-001C-5115	Sequence 5115, Ap
17	32	36.8	177	US-09-134-001C-4482	Sequence 4482, Ap
18	32	36.8	248	US-09-134-001C-5085	Sequence 5085, Ap
19	32	36.8	462	US-09-846-729-3	Sequence 3, Appl1
20	32	36.8	462	US-09-846-729-17	Sequence 17, Appl1
21	32	36.8	464	US-09-846-729-14	Sequence 14, Appl1
22	32	36.8	715	US-09-620-412A-321	Sequence 321, Appl1
23	32	36.8	1693	US-09-851-410-7	Sequence 7, Appl1
24	31	35.6	24	US-09-724-059-445061	Sequence 445061, Ap
25	31	35.6	24	US-09-724-059-447581	Sequence 447581, Ap
26	31	35.6	24	US-09-724-059-449981	Sequence 449981, Ap
27	31	35.6	24	US-09-724-059-452501	Sequence 452501, Ap

28	31	35.6	24	US-09-724-059-457301	Sequence 457301, Ap
29	31	35.6	24	US-09-724-059-462341	Sequence 462341, Ap
30	31	35.6	25	US-09-724-059-572281	Sequence 572281, Ap
31	31	35.6	25	US-09-724-059-577321	Sequence 577321, Ap
32	31	35.6	25	US-09-724-059-582001	Sequence 582001, Ap
33	31	35.6	25	US-09-724-059-584521	Sequence 584521, Ap
34	31	35.6	106	US-09-818-683-364	Sequence 364, App
35	31	35.6	137	US-09-134-001C-3078	Sequence 3078, App
36	31	35.6	218	US-09-497-895-2	Sequence 2, Appl1
37	31	35.6	269	US-09-134-001C-5197	Sequence 5197, App
38	31	35.6	982	US-09-454-684A-176	Sequence 176, App
39	31	35.6	982	US-09-620-412A-176	Sequence 176, App
40	31	35.6	1006	US-09-454-684A-190	Sequence 190, App
41	31	35.6	1006	US-09-620-412A-190	Sequence 190, App
42	30.5	35.1	461	US-09-225-358A-12	Sequence 12, Appl1
43	30.5	35.1	525	US-09-134-001C-3514	Sequence 3514, App
44	30	34.5	24	US-09-724-059-455021	Sequence 455021, Ap
45	30	34.5	24	US-09-724-059-460061	Sequence 460061, Ap

ALIGNMENTS

RESULT 1

US-09-646-043-1

Sequence 1, Application US/09646043

GENERAL INFORMATION:

APPLICANT: HAKALEHTO, Elino

TITLE OF INVENTION: METHOD FOR DETECTING MICROBES FROM AN ENRICHMENT CULTURE

FILE REFERENCE: 0933-0162P

CURRENT APPLICATION NUMBER: US/09/646, 043

CURRENT FILING DATE: 2000-10-25

NUMBER OF SEQ ID NOS: 1

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Peptide sequence derived from Salmonella typhimurium and E.coli

US-09-646-043-1

Query Match 100.0%; Score 87; DB 5; Length 18;

Best Local Similarity 100.0%; Pred. No. 2.4e-08; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASPTAIGDTTAAQVPSIV 18
Db 1 ASPTAIGDTTAAQVPSIV 18

RESULT 2

US-09-134-001C-5569

Sequence 5569, Application US/09134001C

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134, 001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064, 964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055, 779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5569

LENGTH: 557

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5569

```

Query Match      42.5%; Score 37; DB 5; Length 557;
Best Local Similarity 50.0%; Pred. No. 8.8e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0.
OY 3 FTATGDTTAQVP 14
Db 190 FEALGDTSTYEMP 201

```

```

RESULT      3
US-09-134-001C-4010
: Sequence 4010, Application US/09134001C
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4010
: LENGTH: 224
: TYPE: prf
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4010

```

Query Match	41.48;	Score 36;	DB 5;	Length 224;
Best Local Similarity	50.08;	Pred. No. 4.7e+02;		
Matches	7;	Conservative	3;	Mismatches 4;
				Indels 0;
				Gaps 0;

```

RESULT 4
US-09-134-001C-3803
; Sequence 3803, Application US/09134001C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3803
; LENGTH: 293
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3803

```

```

Query Match      41.4%  Score 36: DB 5: Length 293
Best Local Similarity 33.3%  Pred. No. 6, 4e+02:
Matches 6; Conservative 5; Mismatches 7; Indels
QY      1 ASFTTAIGDTTAQPPSIV 18
      |||::||
Db      219 ASFAVSVDGDLNSEPKALI 236

```

```

RESULT      5
US-09-134-001C-4214
: Sequence 4214, Application US/09134001C
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: CTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4214
: LENGTH: 503
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4214

```

Query March	39.1%	Score 34;	DB 5;	Length 503;
Best Local Similarity	40.0%	Pred. No. 2.7e+03;		
Matches	6;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;
QY	4	TAAGDTTAQYPSIV	18	
DB	321	TAAAGDSIVPTNVI	335	

```

RESULT 6
US-09-454-684A-292
; Sequence 292, Application US/09454684A
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhalla, Ajay
; APPLICANT: Skelky, Yasir
; APPLICANT: Filing, Steve
; APPLICANT: Maisonneuve, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.465C4
; CURRENT APPLICATION NUMBER: US/09/454,684A
; CURRENT FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 292
; LENGTH: 333
; TYPE: prt
; ORGANISM: Chlamydia
US-09-454-684A-292

```

Query Match	37.98;	Score 33;	DB 5;	Length 333;
Best Local Similarity	40.08;	Pred. No. 2.5e+03;		
Matches	6;	Conservative	3;	Mismatches 6;
				Indels 0;
				Gaps 0;

```

US-09-620-412A-292
; Sequence 292, Application US/09620412A
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412A
; CURRENT FILING DATE: 2001-07-20

```

NUMBER OF SEQ ID NOS: 357
 SOFTWARE: FastSeq for Windows Version 3.0/4.0
 SEQ ID NO 292
 LENGTH: 333
 TYPE: PRT
 ORGANISM: Chlamydia
 US-09-620-412A-292

Query Match
 Best Local Similarity 37.9%; Score 33; DB 5; Length 333;
 Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 FTAGDTAQPFSIV 18
 DB 167 FAVGATYTSATGVL 161

RESULT 8
 US-09-134-001C-3778
 Sequence 3778, Application US/09134001C
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3778
 LENGTH: 476
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3778

Query Match
 Best Local Similarity 37.9%; Score 33; DB 5; Length 476;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 GDTTAQVPSFIV 18
 DB 294 GDSIATSPFTLV 305

RESULT 9
 US-09-134-001C-5351
 Sequence 5351, Application US/09134001C
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 5351
 LENGTH: 901
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5351

Query Match
 Best Local Similarity 37.9%; Score 33; DB 5; Length 901;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SFATIGDT 9
 DB 467 NFTAIGDS 474

RESULT 10
 US-09-724-059-562321
 Sequence 562321, Application US/09724059
 GENERAL INFORMATION:
 APPLICANT: Choo, Yen
 APPLICANT: Kling, Aaron
 TITLE OF INVENTION: Nucleic Acid Binding Proteins
 FILE REFERENCE: P2500USM
 CURRENT APPLICATION NUMBER: US/09/724,059
 PRIOR FILING DATE: 2000-11-28
 PRIOR APPLICATION NUMBER: GB9710809.6
 PRIOR FILING DATE: 1997-05-23
 NUMBER OF SEQ ID NOS: 1407122
 SOFTWARE: MacAlister
 SEQ ID NO 562321
 LENGTH: 25
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: Polypeptide sequence
 US-09-724-059-562321

Query Match
 Best Local Similarity 36.8%; Score 32; DB 5; Length 25;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 FTAIGDTTA 11
 DB 12 FTRLGDLTA 20

RESULT 11
 US-09-724-059-564841
 Sequence 564841, Application US/09724059
 GENERAL INFORMATION:
 APPLICANT: Choo, Yen
 APPLICANT: Kling, Aaron
 APPLICANT: Isalan, Mark
 TITLE OF INVENTION: Nucleic Acid Binding Proteins
 FILE REFERENCE: P2500USM
 CURRENT APPLICATION NUMBER: US/09/724,059
 PRIOR FILING DATE: 2000-11-28
 PRIOR APPLICATION NUMBER: GB9710809.6
 PRIOR FILING DATE: 1997-05-23
 NUMBER OF SEQ ID NOS: 1407122
 SOFTWARE: MacAlister
 SEQ ID NO 564841
 LENGTH: 25
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: Polypeptide sequence
 US-09-724-059-564841

Query Match
 Best Local Similarity 36.8%; Score 32; DB 5; Length 25;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 FTAIGDTTA 11
 DB 12 FTRLGDLTA 20

RESULT 12
US-09-724-059-567241
; Sequence 567241, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500USM
; CURRENT APPLICATION NUMBER: US/09/724,059
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: Macallister
; SEQ ID NO 567241
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-567241

Query Match 36.8%; Score 32; DB 5; Length 25;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11
11:1111
DB 12 FTQLGDLTA 20

RESULT 13
US-09-724-059-569761
; Sequence 569761, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500USM
; CURRENT APPLICATION NUMBER: US/09/724,059
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: Macallister
; SEQ ID NO 569761
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-569761

Query Match 36.8%; Score 32; DB 5; Length 25;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11
11:1111
DB 12 FTQLGDLTA 20

RESULT 14
US-09-724-059-574561
; Sequence 574561, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark

; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500USM
; CURRENT APPLICATION NUMBER: US/09/724,059
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: Macallister
; SEQ ID NO 574561
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-574561

Query Match 36.8%; Score 32; DB 5; Length 25;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11
11:1111
DB 12 FTQLGDLTA 20

RESULT 15
US-09-724-059-579601
; Sequence 579601, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500USM
; CURRENT APPLICATION NUMBER: US/09/724,059
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: Macallister
; SEQ ID NO 579601
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide sequence
US-09-724-059-579601

Query Match 36.8%; Score 32; DB 5; Length 25;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11
11:1111
DB 12 FTQLGDLTA 20

Search completed: June 20, 2001, 13:42:32
Job time: 164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:38:53 ; Search time 44.73 Seconds

(without alignments)

71.197 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASFTAGDTRAGVPSIV 18

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1316263 seqs, 17692484 residues
Total number of hits satisfying chosen parameters: 1316263

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCRYUS.COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097.COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098.COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	70.1	137	16	US-09-252-691-6209
2	61	70.1	137	16	US-09-252-691-6209
3	47	54.0	171	21	US-09-252-691-6209
4	47	54.0	171	21	US-09-252-691-6209
5	47	54.0	171	21	US-09-252-691-6209
6	47	54.0	171	21	US-09-252-691-6209
7	47	54.0	171	21	US-09-252-691-6209
8	47	54.0	171	21	US-09-252-691-6209
9	47	54.0	171	21	US-09-252-691-6209
10	47	54.0	171	21	US-09-252-691-6209

11	42	48.3	442	19	US-09-540-236-3728	Sequence 3728, Ap
12	42	48.3	442	23	US-60-128-476-4764	Sequence 4764, Ap
13	41	47.1	417	18	US-09-489-039A-11753	Sequence 11753, A
14	41	47.1	504	1	PCT-US00-33349-47	Sequence 47, Appl
15	41	47.1	1101	16	US-09-248-796-17347	Sequence 17347, A
16	40	46.0	152	20	US-09-615-846-13	Sequence 13, Appl
17	40	46.0	388	23	US-60-140-956-1369	Sequence 1369, Ap
18	40	46.0	388	23	US-60-143-992-971	Sequence 971, Ap
19	40	46.0	1732	14	US-09-066-330-11	Sequence 11, Appl
20	40	46.0	1732	14	US-09-174-517-10	Sequence 14, Appl
21	40	46.0	2628	15	US-09-174-517-10	Sequence 40941, A
22	39	44.8	51	16	US-09-270-767-40941	Sequence 56157, A
23	39	44.8	51	16	US-09-270-767-40941	Sequence 194448, A
24	39	44.8	134	1	PCT-US01-01435-108	Sequence 108, Ap
25	39	44.8	164	21	US-09-733-089-3386	Sequence 3386, Ap
26	39	44.8	201	1	PCT-US01-01435-108	Sequence 77, Appl
27	39	44.8	201	1	PCT-US01-01435-108	Sequence 1606, Ap
28	39	44.8	201	1	PCT-US01-01435-108	Sequence 1606, Ap
29	39	44.8	201	1	PCT-US01-01435-108	Sequence 20, Appl
30	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
31	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
32	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
33	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
34	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
35	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
36	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
37	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
38	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
39	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
40	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
41	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
42	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
43	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
44	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
45	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-09-252-691-6209

Sequence 6209, Application US/09252691B

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstein et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252.691B

CURRENT FILING DATE: 1999-02-18

NUMBER OF SEQ ID NOS: 11324

SEQ ID NO 6209

LENGTH: 137

TYPE: PRT

ORGANISM: Enterobacter cloacae

US-09-252-691-6209

Query Match 70.1%; Score 61; DB 16; Length 137;

Best Local Similarity 58.8%; Pred. No. 0.0054;

Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASFTAGDTRAGVPSIV 17

DB 17 AKFTKVDITSNIPFI 33

RESULT 2

US-09-252-691C-6209

Sequence 6209, Application US/09252691C

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstein et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE

```

; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT FILING DATE: 1999-02-18
; CURRENT APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 6209
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-09-252-691C-6209

Query Match          70.1%; Score 61; DB 16; Length 137;
Best Local Similarity 58.8%; Pred. No. 0.0054;
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAYPFSL 17
   |||:||||:~||:~|
Db 17 AKFTKVGDTSNIPFI 33

RESULT 3
US-09-711-164-359
; Sequence 359, Application US/09711164
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlson, Karl
; APPLICANT: Judson, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA-008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-359

Query Match          54.0%; Score 47; DB 21; Length 171;
Best Local Similarity 81.8%; Pred. No. 2.5;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 GDTTAYPFSL 17
   |||||:||||:
Db 65 GDTTAYPFSL 75

RESULT 4
US-60-164-415-359
; Sequence 359, Application US/60164415
; GENERAL INFORMATION:
; APPLICANT: Bauer, Paula
; APPLICANT: Forsyth, Allyn
; APPLICANT: Froelich, Jamie
; APPLICANT: McCarthy, Melissa
; APPLICANT: Ohlson, Karl
; APPLICANT: Phuong, Trung
; APPLICANT: Robbins, David
; APPLICANT: Tan, Zehui
; APPLICANT: Trawick, John
; APPLICANT: Turcek, Bobbi
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA-008PR

; CURRENT APPLICATION NUMBER: US/60/164,415
; CURRENT FILING DATE: 1999-11-09
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-164-415-359

Query Match          54.0%; Score 47; DB 23; Length 171;
Best Local Similarity 81.8%; Pred. No. 2.5;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 GDTTAYPFSL 17
   |||||:||||:
Db 65 GDTTAYPFSL 75

RESULT 5
US-09-489-039A-13057
; Sequence 13057, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13057
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13057

Query Match          54.0%; Score 47; DB 18; Length 173;
Best Local Similarity 52.9%; Pred. No. 2.5;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAYPFSL 17
   |||:||||:~||:~|
Db 59 AOEFTVGTTLPPFPAI 75

RESULT 6
US-09-489-039A-12007
; Sequence 12007, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12007
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12007

Query Match          54.0%; Score 47; DB 18; Length 212;
Best Local Similarity 62.5%; Pred. No. 3.2;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```


OY 3 FTAGDTTAQVPSIV 18
 DB 95 FVAGATTPVPRIV 110

RESULT 7
 US-09-733-089-16780
 ; Sequence 16780, Application US/09733089
 ; GENERAL INFORMATION:
 ; APPLICANT: Dotson, Stanton B.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Lutfiyya, Linda L.
 ; APPLICANT: McIninch, James
 ; APPLICANT: Wu, Wei
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 ; FILE REFERENCE: 38-21(1500)D
 ; CURRENT APPLICATION NUMBER: US/09/733,089
 ; PRIOR FILING DATE: 2000-12-11
 ; PRIOR APPLICATION NUMBER: US 09/474,435
 ; PRIOR FILING DATE: 1999-12-28
 ; PRIOR APPLICATION NUMBER: US 09/654,617
 ; PRIOR FILING DATE: 2000-09-05
 ; PRIOR APPLICATION NUMBER: US 09/620,392
 ; PRIOR FILING DATE: 2000-07-19
 ; NUMBER OF SEQ ID NOS: 24143
 ; SEQ ID NO 16780
 ; LENGTH: 348
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-09-733-089-16780

Query Match 52.9%; Score 46; DB 21; Length 348;
 Best Local Similarity 57.1%; Pred. No. 8.7;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 OY 4 TAIGDTTAQVPSI 17
 DB 293 TVISDLTADIPFSV 306

RESULT 8
 US-09-248-796-19640
 ; Sequence 19640, Application US/09248796
 ; GENERAL INFORMATION:
 ; APPLICANT: Ketch Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796
 ; CURRENT FILING DATE: 1999-02-12
 ; NUMBER OF SEQ ID NOS: 28206
 ; SEQ ID NO 19640
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-248-796-19640

Query Match 49.4%; Score 43; DB 16; Length 437;
 Best Local Similarity 42.9%; Pred. No. 40;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 OY 4 TAIGDTTAQVPSI 17
 DB 128 SSVGDITTLQPFQV 141

RESULT 9
 US-60-238-332-2
 ; Sequence 2, Application US/60238332

GENERAL INFORMATION:
 ; APPLICANT: Lee, Ernestine A.
 ; APPLICANT: Rankumar, Jaya
 ; APPLICANT: Yao, Monique G.
 ; APPLICANT: Patterson, Chandra
 ; APPLICANT: Walla, Narinder K.
 ; APPLICANT: Lu, Yan
 ; APPLICANT: Tribouley, Catherine M.
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Wang, Yumel E.
 ; APPLICANT: Gandhi, Ameena R.
 ; APPLICANT: Hafalia, April
 ; APPLICANT: Xu, Yuming
 ; APPLICANT: Thangavelu, Kavitha
 ; APPLICANT: Baughn, Mariah R.
 ; TITLE OF INVENTION: PROTEIN PHOSPHATASES
 ; FILE REFERENCE: PI-0248 P
 ; CURRENT APPLICATION NUMBER: US/60/238,332
 ; CURRENT FILING DATE: 2000-10-05
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 2
 ; LENGTH: 986
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte ID No: 511906CD1
 US-60-238-332-2

Query Match 49.4%; Score 43; DB 23; Length 986;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 OY 2 SFTAGDTTAQVPSI 17
 DB 74 TETAGADSPQIPFSL 89

RESULT 10
 US-08-277-231-5
 ; Sequence 5, Application US/08277231
 ; GENERAL INFORMATION:
 ; APPLICANT: Green, Bruce A.
 ; APPLICANT: Brinton, Charles C.
 ; TITLE OF INVENTION: Sequence and Analysis of LKP P111n
 ; TITLE OF INVENTION: Structural Genes and The LKP P111 Operon of Nontypable
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Millitia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/277,231
 ; FILING DATE: 19-JUL-1994
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brook, David E.
 ; REGISTRATION NUMBER: 22,592
 ; REFERENCE/DOCKET NUMBER: ACC94-02
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 861-6240
 ; TELEFAX: (617) 861-9540

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 205 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-277-231-5

Query Match 48.3%; Score 42; DB 6; Length 205;
 Best Local Similarity 52.9%; Pred. No. 25;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASETAIGDTTAQVPFSI 17
 DB 96 ANLSHAGOTAAVPFSI 112

RESULT 11
 US-09-540-236-3728
 ; Sequence 3728, Application US/09540236
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
 ; FILE REFERENCE: 2709.2005-001
 ; CURRENT APPLICATION NUMBER: US/09/540,236
 ; CURRENT FILING DATE: 2000-04-04
 ; NUMBER OF SEQ ID NOS: 3840
 ; SEQ ID NO 3728
 ; LENGTH: 442
 ; TYPE: PRT
 ; ORGANISM: M. catarrhalis
 ; US-09-540-236-3728

Query Match 48.3%; Score 42; DB 19; Length 442;
 Best Local Similarity 57.1%; Pred. No. 61;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 FTAIGDTTAQVPFS 16
 DB 95 FTVIGSCAGPFS 108

RESULT 12
 US-60-128-476-4764
 ; Sequence 4764, Application US/60128476
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA
 ; FILE REFERENCE: 107196.139
 ; CURRENT APPLICATION NUMBER: US/60/128,476
 ; CURRENT FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 5002
 ; SEQ ID NO 4764
 ; LENGTH: 442
 ; TYPE: PRT
 ; ORGANISM: Moraxella catarrhalis
 ; US-60-128-476-4764

Query Match 48.3%; Score 42; DB 23; Length 442;
 Best Local Similarity 57.1%; Pred. No. 61;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 FTAIGDTTAQVPFS 16
 DB 95 FTVIGSCAGPFS 108

RESULT 13

US-09-489-039A-11753
 ; Sequence 11753, Application US/09489039A
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 11753
 ; LENGTH: 417
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-11753

Query Match 47.1%; Score 41; DB 18; Length 417;
 Best Local Similarity 60.0%; Pred. No. 87;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 FTAIGDTTAQVPFSI 17
 DB 351 FTAAYVTIARIPEFSI 365

RESULT 14
 PCT-US00-33549-47
 ; Sequence 47, Application PCT/TUS0033549
 ; GENERAL INFORMATION:
 ; APPLICANT: Sette, Alessandro
 ; APPLICANT: Sidney, John
 ; APPLICANT: Southwood, Scott
 ; APPLICANT: Chesnut, Robert
 ; APPLICANT: Cells, Esteban
 ; APPLICANT: Grey, Howard M.
 ; APPLICANT: EpiImmune Inc.
 ; TITLE OF INVENTION: Inducing Cellular Immune Responses to Human
 ; FILE REFERENCE: Compositions
 ; CURRENT APPLICATION NUMBER: PCT/US00/33549
 ; CURRENT FILING DATE: 2000-12-11
 ; PRIOR APPLICATION NUMBER: US 60/172,705
 ; PRIOR FILING DATE: 1999-12-10
 ; PRIOR APPLICATION NUMBER: US 09/641,528
 ; PRIOR FILING DATE: 2000-08-15
 ; NUMBER OF SEQ ID NOS: 86
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 47
 ; LENGTH: 504
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 31
 ; PCT-US00-33549-47

Query Match 47.1%; Score 41; DB 1; Length 504;
 Best Local Similarity 53.3%; Pred. No. 11402;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 FTAIGDTTAQVPFSI 17
 DB 211 FTAIDPTKSNVPLDI 225

RESULT 15
 US-09-248-796-17347
 ; Sequence 17347, Application US/09248796
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI

Thu Jun 21 08:38:32 2001

us-09-646-043-1.rapm

Page 5

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17347
; LENGTH: 1101
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-17347

Query Match 47.18; Score 41; DB 16; Length 1101;
Best Local Similarity 60.08; Pred. No. 2.7e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Oy 1 ASFTAGDGTTOVVF 15
:||||:| ||| |
Db 993 SSFTAVTDFRQVDF 1007

Search completed: June 20, 2001, 13:41:43
Job time: 170 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 18.33 Seconds

(Without alignments)
56.134 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASFTALGDTTACVPSIV 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0401.*

1: /SID6/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID6/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	18	Y29989	Salmonella type 1
2	82	94.3	185	W23584	Salmonella enterit
3	42	48.3	205	R89328	Minor tip-associat
4	42	48.3	1120	Y19780	B. burgdorferi ant
5	42	48.3	1146	Y19779	B. burgdorferi ant
6	41	47.1	215	B16004	E. coli proliferat
7	40	46.0	176	B16007	E. coli proliferat
8	40	46.0	1732	R96029	P. gingivalis porp
9	40	46.0	1732	W24787	Prk antigenic pro
10	40	46.0	1732	W69487	Haemagglutinin pro
11	40	46.0	2628	R96030	P. gingivalis haem

12	40	46.0	2628	19	W69488	Haemagglutinin pro
13	39	44.8	248	21	B41673	Human ORF1437
14	39	44.8	332	21	B34831	Gene 27 human secr
15	39	44.8	393	21	B27624	Human secreted pro
16	39	44.8	393	22	B60757	Gene 12 related pe
17	39	44.8	516	18	W17974	Coplinus clineus
18	39	44.8	516	16	R84825	Modified melanocyt
19	38	43.7	303	21	G45702	Arabidopsis thalia
20	38	43.7	327	21	G45701	Arabidopsis thalia
21	38	43.7	467	21	G45700	Arabidopsis thalia
22	38	43.7	646	18	W13470	Arabidopsis thalia
23	38	43.7	1135	15	R50037	Arabidopsis thalia
24	37	42.5	46	21	G25978	Arabidopsis thalia
25	37	42.5	117	21	Y74376	Arabidopsis thalia
26	37	42.5	118	22	B63602	Arabidopsis thalia
27	37	42.5	226	21	B53207	Arabidopsis thalia
28	37	42.5	259	21	B26244	Arabidopsis thalia
29	37	42.5	1876	16	R72853	Arabidopsis thalia
30	37	42.5	1876	16	R72854	Arabidopsis thalia
31	37	42.5	1895	16	R72854	Arabidopsis thalia
32	36	41.4	106	20	Y25639	Arabidopsis thalia
33	36	41.4	132	20	Y25641	Arabidopsis thalia
34	36	41.4	138	20	Y25636	Arabidopsis thalia
35	36	41.4	138	20	Y25640	Arabidopsis thalia
36	36	41.4	138	20	Y25640	Arabidopsis thalia
37	36	41.4	207	21	B44594	Arabidopsis thalia
38	36	41.4	255	20	Y34658	Arabidopsis thalia
39	36	41.4	278	21	G06910	Arabidopsis thalia
40	36	41.4	292	21	G40208	Arabidopsis thalia
41	36	41.4	535	21	G29571	Arabidopsis thalia
42	36	41.4	546	21	G29570	Arabidopsis thalia
43	36	41.4	681	17	R92701	Arabidopsis thalia
44	36	41.4	702	21	G29569	Arabidopsis thalia
45	36	41.4	830	16	B53128	Macaca mulatta rha
					R65216	p-selectin. Homo

ALIGNMENTS

RESULT 1	Y29989	standard; peptide: 18 AA.
ID	Y29989	
XX		
AC	Y29989;	
XX		
DT	24-NOV-1999	(first entry)
XX		
DE	Salmonella type 1 fimbriae peptide.	
XX		
KW	Salmonella; type 1 fimbriae; immunisation; detection; microbe;	
KW	enteric bacterium; microbiological.	
XX		
OS	Synthetic.	
OS	Salmonella typhimurium.	
OS	Escherichia coli.	
XX		
PN	W09947931-AL.	
XX		
PD	23-SEP-1999.	
XX		
PF	15-MAR-1999;	99WO-FI00192.
XX		
PR	13-MAR-1998;	98FI-0000571.
XX		
PA	(HAKA/) HAKALEHTO E E.	
XX		
PI	Hakalehto EE;	
XX		
DR	WPI; 1999-562222/47.	
XX		
PT	Microbiological determination, useful for detecting microbes in	
XX	clinical samples, food and environmental samples	

Applied
Protein

Db 96 anlshagqtaapvpsi 112

```

RESULT 4
ID Y19780 standard; Protein; 1120 AA.
XX Y19780;
AC Y19780;
XX 19-JUL-1999 (first entry)
DT 19-JUL-1999 (first entry)
XX B. burgdorferi antigenic protein, fl2.aa.
DE B. burgdorferi antigenic protein, fl2.aa.
XX Antigenic protein; vaccine; Lyme disease; infection; detection.
KW Borrelia burgdorferi.
XX Borrelia burgdorferi.
OS MO9859071-A1.
XX 30-DEC-1998.
PD 30-DEC-1998.
XX 18-JUN-1998; 98WO-US12718.
PF 18-JUN-1998; 98WO-US12718.
XX 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX Chai GH, Erwin AL, Hanson MS, Lathigra R;
PI Chai GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI: 1999-189980/16.
DR N-PSDB; X61476.
XX New Isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX Claim 12; Page 57; 275pp; English.
PS This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
CC
SQ Sequence 1120 AA.

Query Match 48.3%; Score 42; DB 20; Length 1120;
Best Local Similarity 53.3%; Pred. No. 51;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPSI 17
ID 1: 11 11 11:1
Db 922 fskigtltklvpsl 936

RESULT 5
ID Y19779 standard; Protein; 1146 AA.
XX Y19779;
AC Y19779;
XX 19-JUL-1999 (first entry)
DT 19-JUL-1999 (first entry)
XX B. burgdorferi antigenic protein, fl2.aa.
DE B. burgdorferi antigenic protein, fl2.aa.
XX Antigenic protein; vaccine; Lyme disease; infection; detection.
KW Borrelia burgdorferi.
XX Borrelia burgdorferi.
OS

```

```

XX MO9859071-A1.
PN 30-DEC-1998.
XX 30-DEC-1998.
PD 30-DEC-1998.
XX 18-JUN-1998; 98WO-US12718.
PF 18-JUN-1998; 98WO-US12718.
XX 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX Chai GH, Erwin AL, Hanson MS, Lathigra R;
PI Chai GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI: 1999-189980/16.
DR N-PSDB; X61475.
XX New Isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX Claim 12; Page 57; 275pp; English.
PS This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
CC
SQ Sequence 1146 AA.

Query Match 48.3%; Score 42; DB 20; Length 1146;
Best Local Similarity 53.3%; Pred. No. 53;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPSI 17
ID 1: 11 11 11:1
Db 948 fskigtltklvpsl 962

RESULT 6
ID B16004 standard; Protein; 215 AA.
XX B16004;
AC B16004;
XX 05-OCT-2000 (first entry)
DT 05-OCT-2000 (first entry)
XX E. coli proliferation associated protein sequence SEQ ID NO:362.
DE E. coli proliferation associated protein sequence SEQ ID NO:362.
XX Escherichia coli; E. coli; proliferation; inhibition; screening;
KW antimicrobial; Bacterial growth; antisense therapy; antibacterial.
XX Escherichia coli.
OS WO200044906-A2.
XX WO200044906-A2.
XX 03-AUG-2000.
PD 03-AUG-2000.
XX 27-JAN-2000; 2000WO-US02200.
PF 27-JAN-2000; 2000WO-US02200.
XX 27-JAN-1999; 99US-0117405.
PR 27-JAN-1999; 99US-0117405.
XX (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA PHARM INC.
XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
PI Yamamoto RT, Xu HH;

```

XX WPI: 2000-514822/46.
 DR N-PSDB: A66010.
 XX
 PT Novel polynucleotides and polypeptides associated with microorganism
 proliferation, used to identify inhibitors of bacterial growth and
 proliferation, for use in antisense therapy -
 XX
 PS Claim 11; Page 270; 316pp; English.
 CC A65809 to A65889 and A66058 to A66138 represent nucleotide sequences
 derived from *Escherichia coli* which inhibit *E. coli* proliferation.
 CC A65890 to A66055 and B15886 to B16040 represent nucleotide and protein
 sequences associated with *E. coli* proliferation. A66056 and A66057
 CC represent primers used for sequencing *E. coli* proliferation inhibiting
 CC nucleotide inserts in an example from the present invention. Methods
 CC from the present invention can be used to identify a proliferation-
 CC required gene in a microorganism, by contacting a microorganism with a
 CC proliferation-required gene activity inhibitory nucleic acid identified
 CC in another organism, and determining if inhibition occurs in the second
 CC microorganism. The nucleic acid sequences identified as being required
 CC for bacterial growth and proliferation, can be used for antisense
 CC therapy for killing bacteria.
 CC
 XX Sequence 215 AA:
 SQ
 Query Match 47.1%; Score 41; DB 21; Length 215;
 Best Local Similarity 53.3%; Pred. No. 11;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 3 FTAGDTTAQVPFSI 17
 | : : : : | : : : : |
 Db 100 fhavgdsapvptvi 114
 RESULT 7
 ID B16007 standard; Protein; 176 AA.
 AC B16007;
 XX
 DT 05-OCT-2000 (first entry)
 DE *E. coli* proliferation associated protein sequence SEQ ID NO:365.
 XX
 KW *Escherichia coli*; *E. coli*; proliferation; inhibition; screening;
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.
 XX
 OS *Escherichia coli*.
 PN WO200044906-A2.
 XX
 PD 03-AUG-2000.
 XX
 PF 27-JAN-2000; 2000WO-US02200.
 XX
 PR 27-JAN-1999; 99US-0117405.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr CJ;
 PI Yamamoto RT, Xu HH;
 DR WPI: 2000-514822/46.
 DR N-PSDB: A66013.
 XX
 PT Novel polynucleotides and polypeptides associated with microorganism
 proliferation, used to identify inhibitors of bacterial growth and
 proliferation, for use in antisense therapy -
 XX
 PS Claim 11; Page 273-274; 316pp; English.
 XX

CC A65809 to A65889 and A66058 to A66138 represent nucleotide sequences
 CC derived from *Escherichia coli* which inhibit *E. coli* proliferation.
 CC A65890 to A66055 and B15886 to B16040 represent nucleotide and protein
 CC sequences associated with *E. coli* proliferation. A66056 and A66057
 CC represent primers used for sequencing *E. coli* proliferation inhibiting
 CC nucleotide inserts in an example from the present invention. Methods
 CC from the present invention can be used to identify a proliferation-
 CC required gene in a microorganism, by contacting a microorganism with a
 CC proliferation-required gene activity inhibitory nucleic acid identified
 CC in another organism, and determining if inhibition occurs in the second
 CC microorganism. The nucleic acid sequences identified as being required
 CC for bacterial growth and proliferation, can be used for antisense
 CC therapy for killing bacteria.
 CC
 XX Sequence 176 AA;
 SQ
 Query Match 46.0%; Score 40; DB 21; Length 176;
 Best Local Similarity 56.2%; Pred. No. 13;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 3 FTAGDTTAQVPFSIV 18
 | : : : : | : : : : |
 Db 59 fmgattpvvpfiri 74
 RESULT 8
 ID R96029 standard; Protein; 1732 AA.
 AC R96029;
 XX
 DT 04-SEP-1996 (first entry)
 DE *P. gingivalis* porphylin.
 XX
 KW Porphylin; haemagglutinin; periodontal disease; vaccine; antibody.
 XX
 OS Porphyromonas gingivalis strain W12.
 PN
 XX
 PD Key Location/Qualifiers
 FT Region 688..708 repeat region type 1"
 FT /note="Pro-Asn repeat region type 1"
 FT 887..952
 FT /note="Pro-Asn repeat region type 2"
 FT 946..967
 FT /note="Pro-Asn repeat region type 1"
 FT 985..1006
 FT /note="Pro-Asn repeat region type 3"
 FT 1041..1100
 FT /note="Pro-Asn repeat region type 4"
 FT 1341..1405
 FT /note="Pro-Asn repeat region type 2"
 FT 1430..1451
 FT /note="Pro-Asn repeat region type 3"
 FT 1488..1547
 FT /note="Pro-Asn repeat region type 4"
 FT 1607..1650
 FT /note="Pro-Asn repeat region type 2"
 FT Region
 XX
 PN WO9617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US16108.
 XX
 PR 09-DEC-1994; 94US-0353485.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Han N, Lantz M, Lepine G, Patil JM, Progniske-Fox A;
 PI Tumwasorn S;
 PI

PA (UYFL) UNIV FLORIDA.
 XX Han N, Lantz M, Lepine G, Patti JM, Proguliske-Fox A;
 PI Tumwasorn S;
 XX WPI: 1998-582627/49.
 DR N-PSDB: V58874.
 XX
 PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin
 and/or protease poly:peptide(s))
 PS Claim 1; Column 69-84; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the prp haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease.
 XX
 SQ Sequence 1732 AA;

Query Match 46.0%; Score 40; DB 19; Length 1732;
 Best Local Similarity 37.5%; Pred. No. 2e+02; 4; Indels 0; Gaps 0;
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SPTAIGDTTAQVPEFI 17
 : : : | | | | : : :
 Db 1658 nitsladvtaqkpytl 1673

RESULT 11
 R96030 ID R96030 standard; Protein; 2628 AA.
 XX
 AC R96030;
 XX
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hgaA.
 XX
 KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
 XX
 OS Porphyromonas gingivalis strain 381.
 XX
 FH Key Location/Qualifiers
 FT Peptide 5..21
 FT /label= Sig-peptide
 XX
 FM W09617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US16108.
 XX
 PR 09-DEC-1994; 94US-0353485.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Han N, Lantz M, Lepine G, Patti JM, Proguliske-Fox A;
 PI Tumwasorn S;
 XX
 DR WPI: 1996-287181/29.
 DR N-PSDB: T30654.
 XX
 PT Porphyromonas gingivalis genes and proteins - used in the detection
 PT and vaccination against periodontal disease
 XX
 PS Claim 6; Page 93-101; 153pp; English.
 XX
 CC P. gingivalis 381 haemagglutinin hgaA (R96030) was identified as

CC the product of the hgaA gene (T30654) isolated as an ECoRV
 CC fragment of genomic DNA. The haemagglutinin, or portions of it
 CC (see also R96021-24), can be obt. from transformed host cells and
 CC used as a vaccine to protect humans or animals against periodontal
 CC disease. Expression in Salmonella cells allows produ. of a live
 CC vaccine. The haemagglutinin can also be used to detect the
 CC presence of anti-P. gingivalis antibodies and to raise monoclonal
 CC antibodies for diagnostic appln.
 XX
 SQ Sequence 2628 AA;

Query Match 46.0%; Score 40; DB 17; Length 2628;
 Best Local Similarity 37.5%; Pred. No. 3.3e+02;
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SPTAIGDTTAQVPEFI 17
 : : : | | | | : : :
 Db 2554 nitsladvtaqkpytl 2569

RESULT 12
 W69488 ID W69488 standard; Protein; 2628 AA.
 XX
 AC W69488;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX
 OS Porphyromonas gingivalis.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 PE 11-DEC-1995; 95US-0570311.
 XX
 PR 11-DEC-1995; 95US-0570311.
 PR 08-SEP-1988; 88US-0241640.
 PR 25-JAN-1991; 91US-0647119.
 PR 09-DEC-1994; 94US-0353485.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Han N, Lantz M, Lepine G, Patti JM, Proguliske-Fox A;
 PI Tumwasorn S;
 XX
 DR WPI: 1998-582627/49.
 DR N-PSDB: V58875.
 XX
 PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin
 PT and/or protease poly:peptide(s))
 PS Claim 1; Column 91-110; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease.
 XX
 SQ Sequence 2628 AA;

Query Match 46.0%; Score 40; DB 19; Length 2628;
 Best Local Similarity 37.5%; Pred. No. 3.3e+02;
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 SFTAGDTTAYVPSIV 17
 ID B41673 standard; Protein; 248 AA.
 AC B41673;
 XX 08-FEB-2001 (first entry)
 DT Human ORFX ORF1437 polypeptide sequence SEQ ID NO:2874.
 DE
 XX Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 XX MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 XX 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; C75882.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11; Page 2112; 5507pp: English.

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 248 AA;
 QY 5 AIGDTTAYVPSIV 18
 ID B34831 standard; Protein; 332 AA.
 AC B34831;
 XX 26-JAN-2001 (first entry)
 DT Human; secreted protein homologous amino acid sequence #119.
 DE
 XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
 KW cancer; immune disorder; cardiovascular disorder; wound healing;
 KW neurological disease; infectious disease; chromosome identification.
 XX
 OS Synchocystis sp.
 XX
 XX MO200058356-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 22-MAR-2000; 2000MO-US07535.
 XX
 XX 26-MAR-1999; 9905-0126511.
 PR 17-DEC-1999; 9905-0172413.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI; 2000-594639/56.
 XX
 XX Fifty nucleic acid molecules encoding human secreted proteins, useful
 PT in the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases -
 XX
 XX Disclosure; Page 404-405; 425pp: English.

CC The polynucleotide sequences given in C59966 to C60015 encode the human
 CC secreted proteins given in B34773 to B34822. B34823 to B34852 represent
 CC human secreted polypeptide sequences and proteins homologous to them,
 CC which are given in the exemplification of the present invention. Human
 CC secreted proteins have activities based on the tissues and cells the
 CC genes are expressed in. Examples of activities include: cytostatic;
 CC immunosuppressive; neurotropic; neuroprotective; antiviral; antiallergic;
 CC hepatotropic; antidiabetic; antineoplastic; antiparasitic; and cardiant;
 CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
 CC The polynucleotides and polypeptides are useful for preventing, treating
 CC or ameliorating a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. The polypeptides can also
 CC be used as a food additive or preservative to increase or decrease
 CC storage capabilities. The polynucleotides are useful for chromosome
 CC identification. They are also useful as probes for diagnosing a disorder
 CC related to the female reproductive system, particularly breast and/or

CC ovary cancer. They are also useful in the gene therapy of breast and
 CC ovarian cancer. Nucleic acids, protein, antibodies, agonists and
 CC antagonists from the present invention are useful in the diagnosis,
 CC treatment and prevention of cancer, immune disorders, cardiovascular
 CC disorders, wound healing, neurological diseases and infectious disease.
 CC C59957 to C59965 and B34772 represents sequence used in the
 CC exemplification of the present invention.

XX Sequence 332 AA;

Query Match 44.8%; Score 39; DB 21; Length 332;
 Best Local Similarity 38.9%; Pred. No. 42;
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAQVPSIV 18
 1:11 | | :1:1:
 DB 173 atfvngktevltpydli 190

RESULT 15

B27624
 ID B27624 standard; Protein: 393 AA.

AC B27624;

DT 02-FEB-2001 (first entry)

DE Human secreted protein BLAST search protein SEQ ID NO: 125.

XX Cytostatic; immunosuppressive; nocitropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO20005175-A1.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000MO-US06049.

XX 12-MAR-1999; 99US-0124144.

XX 11-JUN-1999; 99US-0138574.

XX 03-DEC-1999; 99US-0168667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638175/61.

XX Isolated nucleic acid molecule encoding a human secreted protein is
 XX used in preventing, treating or ameliorating a medical condition -

XX Disclosure; Page 410-412; 428pp; English.

XX The invention relates to the isolation of genes C58990-C59039 encoding
 CC the human secreted proteins B27560-B27609. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene given
 CC in the descriptor line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
 CC as myocardial ischaemias; (d) wound healing; (e) neurological diseases
 CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
 CC as viral, bacterial, fungal and parasitic infections.

XX Sequence 393 AA;

Query Match 44.8%; Score 39; DB 21; Length 393;
 Best Local Similarity 38.9%; Pred. No. 52;
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAQVPSIV 18
 1:11 | | :1:1:
 DB 234 atfvngktevltpydli 251

Search completed: June 20, 2001, 13:32:02
 Job time: 40 sec

